10-9-2020

Selling Consumer Data for Profit: Optimal Market-Segmentation Design and its Consequences

Kai Hao Yang

SELLING CONSUMER DATA FOR PROFIT: OPTIMAL MARKET-SEGMENTATION DESIGN AND ITS CONSEQUENCES

By

Kai Hao Yang

October 2020

COWLES FOUNDATION DISCUSSION PAPER NO. 2258

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

Selling Consumer Data for Profit:
Optimal Market-Segmentation Design and its Consequences

Kai Hao Yang*

October 9, 2020

Abstract

A data broker sells *market segmentations* created by consumer data to a producer with private production cost who sells a product to a unit mass of consumers with heterogeneous values. In this setting, I completely characterize the revenue-maximizing mechanisms for the data broker. In particular, every optimal mechanism induces *quasi-perfect price discrimination*. That is, the data broker sells the producer a market segmentation described by a cost-dependent cutoff, such that all the consumers with values above the cutoff end up buying and paying their values while the rest of consumers do not buy. The characterization of optimal mechanisms leads to additional economically relevant implications. I show that the induced market outcomes remain unchanged even if the data broker becomes more active in the product market by gaining the ability to contract on prices; or by becoming an exclusive retailer, who purchases both the product and the exclusive right to sell the product from the producer, and then sells to the consumers directly. Moreover, vertical integration between the data broker and the producer increases total surplus while leaving the consumer surplus unchanged, since consumer surplus is zero under any optimal mechanism for the data broker.

KEYWORDS: Price discrimination, market segmentation, mechanism design, virtual cost, quasi-perfect segmentation, quasi-perfect price discrimination, surplus extraction, outcome-equivalence

JEL CLASSIFICATION:D42, D82, D61, D83, L12

*Cowles Foundation for Economic Research, Yale University, kaihao.yang@yale.edu. I am indebted to my advisor Phil Reny for his constant support and encouragement, and to my thesis committee: Ben Brooks, Emir Kamenica and Doron Ravid for their invaluable guidance and advices. I appreciate the helpful comments and suggestions from Mohammad Akborpour, Dirk Bergemann, Isa Chaves, Yeon-Koo Che, Alex Frankel, Andrew Gianou, Andreas Kleiner, Jacob Leshno, Elliot Lipnowski, Alejandro Maelli, Roger Myerson, Barry Nalebuff, Michael Ostrovsky, Marek Pycia, Daniel Rappoport, Eric Rasmusen, Ilya Segal, Andy Skrzypacz, Wenji Xu and Weijie Zhong. I also thank the participants of various conferences and seminars at which this paper was presented. All errors are my own.

1 Introduction

1.1 Motivation

The rapid development of informational technology has enabled one to collect, process and analyze vast volumes of consumer data. By the use of consumer data, the scope of price discrimination has moved far beyond its traditional boundaries such as geography, age, or gender. Extensive usage of consumer data allows one to identify many characteristics of consumers that are relevant to the prediction of their values, and therefore to create numerous sorts of *market segmentation*—a way to split the market demand into several segments by partitioning the consumers' characteristics—to facilitate price discrimination. Moreover, because of their specialization in information technology, several "data brokers" trade vast amounts of consumer data with retailers, which effectively means these data brokers can create market segmentations and sell them as a product that facilitates price discrimination. For instance, online platforms such as Facebook collect and sell[1] a significant amount of consumers' personal information, including personal characteristics, traveling plans, lifestyles, and text messages via its own platform. Alternatively, data companies such as Acxiom and Datalogix gather and sell personal information such as government records, financial activities, online activities and medical records to retailers (Federal Trade Commission, 2014).

This paper studies the design of optimal selling mechanisms of a data broker. In this paper, I consider a model where there is one producer with privately known constant marginal cost, who produces and sells a single product to a unit mass of consumers. The consumers have unit demand and the distribution of their values is described by commonly known market demand. Into this environment, I introduce a *data broker*, who does not know the producer's marginal cost of production but can sell *any* market segmentation to the producer via *any* selling mechanism. As the data broker is uncertain about the production cost, and only affects the product market indirectly by selling consumer data to the producer, it is not obvious how the data broker should sell market segmentations to the producer, what market segmentations will be created, and how the sale of consumer data affects economic welfare and allocative outcomes.

The main result of this paper is a complete characterization of the revenue-maximizing mechanisms for the data broker. The optimal mechanisms feature *quasi-perfect price discrimination*, which is an outcome where all the purchasing consumers pay exactly their values,

[1] In practice, "selling" consumer data can take a wide variety of forms, which include not only traditional physical transactions but also integrated data-sharing agreements/activities. For instance, in a recent full-scale investigation by *The New York Times*, Facebook has formed ongoing partnerships with other firms, including Netflix, Spotify, Apple and Microsoft, and granted these companies accesses to different aspects of consumer data *"in ways that advanced its own interests."* See full news coverage at https://www.nytimes.com/2018/12/18/technology/facebook-privacy.html

although not every consumer with values above the marginal cost buys the product. Theorem 1 shows that every optimal mechanism must create *quasi-perfect segmentations* described by a cost-dependent cutoff. Namely, all the consumers with values above the cutoff are separated from each other whereas the consumers with values below the cutoff are pooled with the separated high-value consumers. When pricing optimally under this segmentation, the producer only sells to high-value consumers and induces quasi-perfect price discrimination. Furthermore, the cutoff function under any optimal mechanism is exactly the minimum between the (ironed) *virtual marginal cost* function and the optimal uniform price as a function of marginal cost. With proper regularity conditions, Theorem 2 shows that there is an optimal mechanism where conditional on being below the cutoff, the distribution of consumer values is the same as the market demand in every market segment. This optimal mechanism can be implemented by a menu of *price-recommendation* data, in which each item consists of a list of recommended prices, one for each consumer, as well as an associated amount that has to be paid to the data broker for this list. Each item in this menu is described by a cutoff, so that for consumers with values above the cutoff, the recommended prices for them equal to their values. Meanwhile, for the consumers with values below the cutoff, the recommended prices for them are distributed in the same way as the consumers' values conditional on being above the cutoff.

The characterization of the optimal mechanisms further leads to several economically relevant implications. As the defining feature of quasi-perfect price discrimination, under any optimal mechanism, all the consumers pay their values conditional on buying. This implies that the consumer surplus under any optimal mechanism is zero (Theorem 3). In other words, in terms of consumer surplus, it is *as if* all the information about the consumers' values is revealed to the producer. The fact that the data broker only affects the product market indirectly via data provision does not benefit the consumers. Furthermore, Theorem 1 also allows a comparison between data brokership and uniform pricing, where no consumer data can be shared. More specifically, I show that data brokership always increases total surplus (Theorem 4), and can even be Pareto-improving compared with uniform pricing if the data broker has to purchase the data from the consumers (before they learn their values, see Theorem 5).

In addition to the welfare implications, another set of relevant questions pertain to how different market regimes would affect market outcomes. More specifically, how would the market outcomes differ if the data broker, instead of merely providing consumer data to the producers, plays a more active role in the product market? The characterization given by Theorem 1 allows comparisons across several other natural market regimes in addition to data brokership, including *vertical integration*, where the data broker and the producer merge and all the private information about production cost is revealed; *direct acquisition*, where the data broker is able purchase the entire production technology from the producer; *exclusive retail*, where the data broker negotiates with the producer and purchases the product, as well as the exclusive right to sell the product, from the producer; and *price-controlling data brokership*, where the data broker can prescribe what price to charge in addition to providing consumer data. Using the main characterization, I show that vertical integration between the data broker and producer increases total surplus while leaving the consumer surplus unchanged (Theorem 6). Furthermore, in terms of market outcomes (i.e., data broker's revenue, producer's profit, consumer surplus and the allocation of the product), I show that data brokership is equivalent to both exclusive retail and price-controlling data brokership (Theorem 7). Together with the closed-form characterization of the data broker's optimal selling mechanism, it can be calculated whether direct acquisition is more profitable than the other market regimes.

The rest of this paper is organized as follows. Continuing in this section, several related literatures are discussed. In Section 2, I provide an illustrative example to demonstrate the design of the data broker's optimal selling mechanism and its implications. The preliminaries and the model are presented in Section 3. In Section 4, I characterize the optimal mechanisms of the data broker. Consequences of consumer-data brokership are discussed in Section 5. Finally, relaxations of the assumptions and several extensions can be found in Section 6 and some further discussions are in Section 7.

1.2 Related Literature

This paper is related to various streams of literature. In the literature of price discrimination, several theoretical works center around the welfare effects of price discrimination (see, for instance, Varian (1985), Aguirre et al. (2010) and Cowan (2016)) and provide conditions under which third-degree price discrimination increases or decreases total surplus and output. In addition, Bergemann et al. (2015) show that any surplus division between the consumers and a monopolist can be achieved by some market segmentation.[2] In these papers, market segmentation is treated as an exogenous object, whereas in my paper, market segmentation is determined endogenously by a data broker's revenue-maximization behavior. Furthermore, Wei and Green (2020) study another channel of price discrimination that does not involve market segmentation (i.e., through providing differential information).

Furthermore, this paper is related to the recent literature of the sale of information by a monopolistic information intermediary. Admati and Pfleiderer (1985) and Admati and Pfleiderer (1990) consider a monopoly who sells information about an asset in a speculative market. Bergemann and Bonatti (2015) explore a pricing problem of a data provider who

[2]See also: Haghpanah and Siegel (2020), who further consider segmentations in environments that feature second-degree price discrimination.

provides data to facilitate targeted marketing. Bergemann et al. (2018) solve a mechanism design problem in which the designer sells experiments to a decision maker who has private information about his belief. Yang (2019) considers a model where an intermediary can provide information about the product to the consumers and charge the seller for such services. Segura-Rodriguez (2020) studies the revenue maximization of a data broker who sells data to firms that differ in the consumer characteristics they wish to forecast.[3]

Methodologically, this paper is related to the literature of mechanism design and information design (see, for instance, Mussa and Rosen (1978), Myerson (1981), Kamenica and Gentzkow (2011) and Bergemann and Morris (2016)). More particularly, my paper can be regarded as a mechanism design problem where the information structure is also part of the design object (see, for instance, Bergemann and Pesendorfer (2007), Yamashita (2017) and Dworczak (2020)).

Among the aforementioned papers, Bergemann et al. (2015), Bergemann et al. (2018) and Yang (2019) are the closest to my paper. Specifically, Bergemann et al. (2015) explore the welfare implications of different market segmentations, while I introduce a data broker who designs the market segmentation in order to maximize revenue. Bergemann et al. (2018) study an environment where the agent has private information about his prior belief and characterize the optimal mechanism in a binary-action, binary-state case; or in a binary-type case, while my paper studies a revenue maximization problem where the agent's private information is part of her intrinsic preference and has a rich action space. Also, this paper allows a large class of priors, including those with infinite support. Finally, Yang (2019) solves for optimal mechanisms of an intermediary that can provide information about the product to the consumers, while in this paper I consider the case where an intermediary sells information about the consumers' values to the producer.

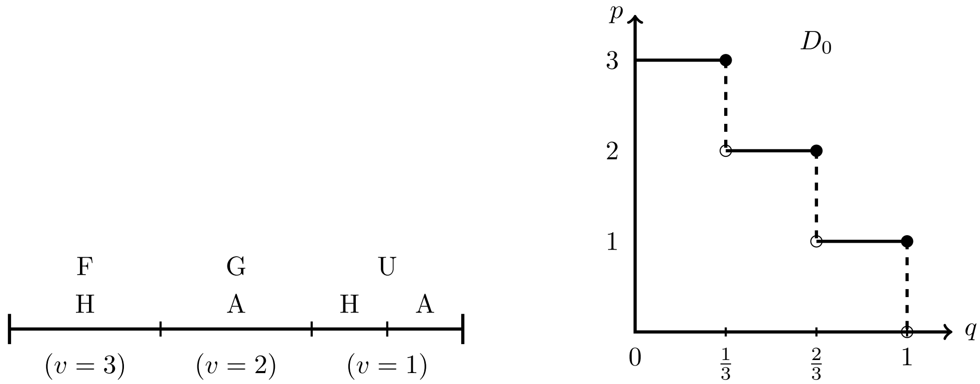
2 An Illustrative Example

To fix ideas, consider the following simple example. A publisher sells an advanced textbook for graduate study. Her (constant) marginal cost of production c is her private information and takes two possible values, $1/4$ or $3/4$, with equal probability. There is a unit mass of consumers with three possible occupations: faculty, undergraduate students, and graduate students. Each of them constitutes $1/3$ of the entire population. It is common knowledge that the textbook has has value $v = 1$ for an undergraduate student, value $v = 2$ for a graduate student and value $v = 3$ for a faculty member. In addition, suppose that among all the undergraduate students, $1/2$ live in houses and $1/2$ live in apartments, whereas all

[3]Relatedly, Acemoglu et al. (2019), Bergemann et al. (2020) and Ichihashi (2020) examine environments where a data broker buys data from the consumers and then sells the consumer data to downstream firms.

the graduate students live in apartments and all the faculty members live in houses. This economy can be represented by Figure 1, where Figure 1a plots the partitions of the consumers induced by their occupations and residence types and Figure 1b plots the (inverse) market demand D_0.

Figure 1: Representation of the market



(a) Partitioning consumers (b) Market demand D_0

Suppose that there is a data broker who owns all the data about the consumers (e.g., income, medical records, occupations and residential information) and thus is able to provide any partition on the line in Figure 1a to the publisher. How should the data broker sell these data to the publisher? An intuitive guess is that the data broker should sell the most informative data. That is, he should provide the publisher with occupation data so that each consumer's value can be fully revealed. Upon receiving such data, the publisher is able to perfectly price discriminate the consumers. The value-revealing data creates a *market segmentation* that decomposes the market into three market segments, and each market segment enables the publisher to perfectly identify the value of the consumers in that market segment. As a result, if the price of the value-revealing data is τ and if the publisher with cost $c \in \{1/4, 3/4\}$ buys the data, her net profit would be

$$\frac{1}{3}(1-c) + \frac{1}{3}(2-c) + \frac{1}{3}(3-c) - \tau.$$

Alternatively, if the publisher with cost c does not buy any data, she would then charge an optimal uniform price (either $1, 2$ or 3, since these are the only possible consumer values) and earn profit

$$\max\left\{(1-c), \frac{2}{3}(2-c), \frac{1}{3}(3-c)\right\}.$$

Therefore, for any τ, the publisher with cost c would buy the value-revealing data if and only if

$$\frac{1}{3}(1-c) + \frac{1}{3}(2-c) + \frac{1}{3}(3-c) - \tau \geq \max\left\{(1-c), \frac{2}{3}(2-c), \frac{1}{3}(3-c)\right\},$$

which simplifies to $\tau \leq (2-c)/3$. Thus, since $c \in \{1/4, 3/4\}$, when $\tau \leq 5/12$, the publisher would always buy the value-revealing data regardless of her marginal cost. When $5/12 < \tau \leq 7/12$, the publisher would buy the data only if $c = 1/4$. Therefore, charging a price $\tau = 5/12$ gives the data broker revenue $5/12$ whereas charging a price $\tau = 7/12$ gives the data broker revenue $7/12 \times 1/2 = 7/24 < 5/12$. Hence the optimal price for the value-revealing data is $5/12$ and it gives the data broker revenue $5/12$.

However, the data broker can in fact improve his revenue by creating a menu consisting of not just the value-revealing data. To see this, consider the following menu of data

$$\mathcal{M}^* = \left\{ \left(\text{residential data}, \tau = \frac{1}{3} \right), \left(\text{value-revealing data}, \tau = \frac{7}{12} \right) \right\}.$$

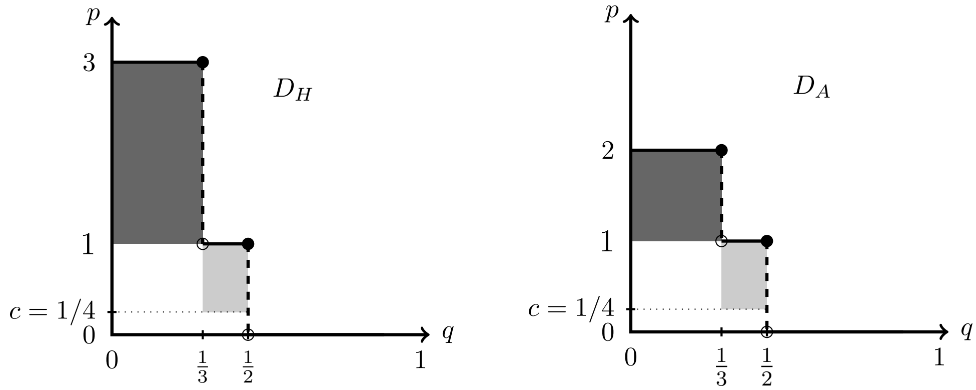
Notice that the residential data creates a market segmentation with two segments described by two demand functions, D_H and D_A. Segment D_H contains all of the consumers with $v = 3$ and $1/2$ of the consumers with $v = 1$ (i.e., those who live in houses), while segment D_A contains all of the consumers with $v = 2$ and $1/2$ of the consumers with $v = 1$ (i.e., those who live in apartments). Figure 2 plots this market segmentation. From Figure 2, it can be seen that $D_H + D_A = D_0$. Moreover, for the publisher with $c = 1/4$, the difference in profit between charging price 3 (2) and charging price 1 in segment D_H (D_A) is exactly the difference between the area of the darker region and the area of the lighter region depicted in Figure 2. Therefore, since the area of the lighter region is smaller than the area of the darker region, charging a price of 3 (2) is better than charging a price of 1 in segment D_H (D_A). Thus, as there are only two possible values in each segment, charging a price of 3 (2) is optimal for the publisher under segment D_H (D_A). This is also the case when her cost is $c = 3/4$, since the area of the lighter region would decrease and the area of the darker region would remain unchanged when the marginal cost changes from $1/4$ to $3/4$. As a result, regardless of her marginal cost, the publisher will sell to all the consumers with values $v = 3$ and $v = 2$ by charging exactly their values upon receiving the residential data.

With this observation, it then follows that when $c = 1/4$, the publisher would prefer buying the value-revealing data (at the price of $\tau = 7/12$) whereas when $c = 3/4$, the publisher would prefer buying the residential data (at the price of $\tau = 1/3$). Therefore, when menu $\mathcal{M}^*$ is provided, the data broker's revenue is

$$(0.5)\frac{1}{3} + (0.5)\frac{7}{12} = \frac{11}{24} > \frac{5}{12},$$

which is higher than what can be obtained by selling value-revealing data alone. The intuition behind such an improvement is simple. When selling the value-revealing data alone, the publisher with lower marginal cost retains more rents because the broker would have to incentivize the high-cost publisher to purchase. However, by creating a menu containing both the value-revealing data and the residential data, the data broker can further screen the

Figure 2: Market segmentation induced by residential data



publisher. To see this, notice that even though the residential data is less informative than the value-revealing data, the only extra benefit of the value-revealing data is for the publisher to be able to price discriminate the consumers with $v = 1$. Thus, when the publisher's marginal cost is high (i.e., $c = 3/4$), the additional information given by the value-revealing data is less useful to the publisher because the gains from selling to consumers with $v = 1$ are small. In contrast, when the publisher has a low marginal cost (i.e., $c = 1/4$), the value-revealing data is more valuable to the publisher since the gains from selling to consumers with $v = 1$ are larger. Therefore, by providing a menu that contains two different datasets with different prices, the data broker can screen the publisher and extract more revenue from the publisher with lower marginal cost than by selling the value-revealing data alone.

In fact, as it will be shown in Section 4, $\mathcal{M}^*$ is an optimal mechanism of the data broker. The optimal mechanism $\mathcal{M}^*$ has several notable features. First, when $c = 3/4$, the high-value consumers ($v = 2$ and $v = 3$) are separated from each other whereas the low-value consumers ($v = 1$) are pooled together with the high-value consumers. This induces a market outcome where consumers with values $v = 2$ and $v = 3$ are buying the textbook by paying their values, whereas the consumers with $v = 1$ do not buy, even if their value is greater than the publisher's marginal cost $3/4$. In other words, in order to maximize revenue, the data broker would sometimes discourage (ex-post) efficient trades. Second, all the purchasing consumers are paying exactly their values, which implies that consumer surplus is zero. Finally, even though every purchasing consumer pays their value, the high-cost publisher never learns exactly each individual consumer's value. These features are not specific to this simple example. In fact, all of them hold in a general class of environments, which will be explored later in this paper.

3 Model

3.1 Notation

The following notation is used throughout the paper. For any Polish space X, $\Delta(X)$ denotes the set of probability measures on X where X is endowed with the Borel σ-algebra and $\Delta(X)$ is endowed with the with weak-* topology. When $X = [\underline{x}, \overline{x}] \subseteq \mathbb{R}$ is an interval, let $\mathcal{D}(X)$ denote the collection of nonincreasing and left-continuous functions $D : X \to [0, 1]$ such that $D(\underline{x}) = 1$, $D(\overline{x}^{+}) = 0$. Since $\mathcal{D}(X)$ and $\Delta(X)$ are bijective, for any $D \in \mathcal{D}(X)$, let $m^D \in \Delta(X)$ be the probability measure associated with D and define the integral

$$\int_A h(x)D(\mathrm{d}x) := \int_A h(x)m^D(\mathrm{d}x),$$

for any measurable $h : X \to \mathbb{R}$. Then, endow $\mathcal{D}(X)$ with the weak-* topology and the Borel σ-algebra using this integral (details in Appendix A). Also, write $\mathrm{supp}(D) := \mathrm{supp}(m^D)$.

3.2 Primitives

There is a single product, a unit mass of consumers with unit demand, a producer for this product (she), and a *data broker* (he). Across the consumers, their values v for the product are distributed according to a commonly known probability measure $m^0 \in \Delta(V)$ and thus can be described by a *market demand* $D_0 \in \mathcal{D} := \mathcal{D}(V)$, where $D_0(p) := m^0([p, \overline{v}])$ is the share of consumers whose values are above p and $V = [\underline{v}, \overline{v}] \subset \mathbb{R}_+$ is a compact interval. Each consumer knows their own value. For the rest of the paper, D_0 is said to be *regular* if the marginal revenue function induced by D_0 is decreasing.[4]

The producer has a constant marginal cost of production $c \in C = [\underline{c}, \overline{c}] \subset \mathbb{R}_+$ for some $0 \leq \underline{c} < \overline{c} < \infty$. The marginal cost c is private information to the producer and follows a distribution G, where G has a density $g > 0$ and induces a *virtual (marginal) cost* function ϕ_G, defined as $\phi_G(c) := c + G(c)/g(c)$ for all $c \in C$. Henceforth, G is said to be *regular* if ϕ_G is increasing.

The data broker owns all the consumer data and can create any *market segmentation*, which is a probability measure $s \in \Delta(\mathcal{D})$ that satisfies the following condition

$$\int_{\mathcal{D}} D(p)s(\mathrm{d}D) = D_0(p),\ \forall p \in V. \tag{1}$$

That is, a segmentation is a way to split the market demand D_0 into different *market segments*.[5] Let $\mathcal{S}$ denote the set of segmentations.

[4] More specifically, D_0 is regular if the function $p \mapsto pD_0(p)$ is single-peaked on $\mathrm{supp}(D_0)$.

[5] As illustrated in the motivating example, different consumer data induce different partitions of consumers' characteristics and thus different ways to split D_0 into a collection of demand functions that sum up to

3.3 Timing of the Events

First, the data broker proposes a *mechanism*, which contains a set of available messages that the producer can send, as well as mappings that specify the market segmentation and the amount of transfers as functions of the messages. Then, the producer decides whether to participate in the mechanism. If the producer does not, she only operates under D_0 without any further segmentations and pays nothing. If the producer participates in the mechanism, she sends a message from the message space, pays the associated transfer, and then operates under the associated market segmentation.

Given any segmentation $s \in \mathcal{S}$, the producer engages in price discrimination by choosing a price $p_D \geq 0$ in each segment $D \in \text{supp}(s)$.[6] To maximize profit, for any segment $D \in \text{supp}(s)$, the producer with marginal cost c solves

$$\max_{p \in \mathbb{R}_+} (p-c)D(p).$$

For any $c \in C$ and any $D \in \mathcal{D}$, let $\boldsymbol{P}_D(c)$ denote the set of optimal prices for the producer with marginal cost c under market segment D. As a convention, regard $\boldsymbol{P}$ as a correspondence on $\mathcal{D} \times C$ and if $\boldsymbol{p}$ is a selection for $\boldsymbol{P}$, write $\boldsymbol{p} \in \boldsymbol{P}$.[7] Furthermore, for any $c \in C$ and any $D \in \mathcal{D}$, let

$$\pi_D(c) := \max_{p \in \mathbb{R}_+} (p-c)D(p)$$

denote the maximized profit of the producer. Also, let

$$\overline{\boldsymbol{p}}_D(c) := \max \boldsymbol{P}_D(c)$$

D_0. Thus, given a market segmentation s, each market segment $D \in \text{supp}(s)$ can be interpreted as a group of consumers who have some common characteristics (e.g., house residents). Notice that by allowing the data broker to provide *any* market segmentation, it is implicitly assumed that the data broker always has enough to data to identify each consumer's value and is able to segment the consumers according to their values arbitrarily. In Section 6, I further discuss an extension that permits the data broker to have incomplete information about the consumers' values. In Section 7, I argue that as long as the consumer characteristics are "rich enough", this definition of market segmentation is equivalent to partitioning the consumer characteristics.

[6] It is without loss of generality to restrict attention to posted price mechanisms even though the producer has private information about c when designing selling mechanisms. This is because the environment features independent private values and quasi-linear payoffs, and both the producer's and the consumers' payoffs are monotone in their types. By Proposition 8 of Mylovanov and Tröger (2014), it is as if c is commonly known when the producer designs selling mechanisms. Therefore, since the consumers have unit demand, according to Myerson (1981) and Riley and Zeckhauser (1983), it is without loss to restrict attention to posted price mechanisms.

[7] For notational conveniences, I restrict the feasible prices for each producer to a large enough compact interval $\overline{V} \subset \mathbb{R}_+$ such that $V \subsetneq \overline{V}$. With this restriction, $\boldsymbol{P}_D(c)$ would be a subset of a compact interval for all $D \in \mathcal{D}$ and for all $c \in C$. Since V itself is bounded, this restriction is simply a notational convention and does not affect the model at all.

be the largest optimal price for the producer with marginal cost c under market segment D.[8] For conciseness, let $\overline{\boldsymbol{p}}_0 := \overline{\boldsymbol{p}}_{D_0}$.

Throughout Section 4 and Section 5, I impose the following technical assumption on the market demand D_0 and the distribution G.

Assumption 1. *The function $c \mapsto \max\{g(c)(\phi_G(c) - \overline{\boldsymbol{p}}_0(c)), 0\}$ is nondecreasing.*

Assumption 1 permits a wide class of (D_0, G) and includes many common examples.[9] Also, it does not require regularities of either D_0 or G (nor is it implied by regularities of D_0 and G). In Section 6, I will further discuss this assumption, including how the results rely on it, its relaxations, as well as several economically interpretable sufficient conditions.

3.4 Mechanism

When proposing mechanisms, by the revelation principle (Myerson, 1979), it is without loss to restrict the data broker's choice of mechanisms to incentive compatible and individually rational *direct mechanisms* that ask the producer to report her marginal cost and then provide the segmentation and determine the transfer accordingly.[10]

Formally, a mechanism is a pair (σ, τ), where $\sigma : C \to \mathcal{S}$, $\tau : C \to \mathbb{R}$ are measurable functions. Given a mechanism (σ, τ), for each $c \in C$, $\sigma(c) \in \mathcal{S}$ stands for the market segmentation provided to the producer, $\tau(c) \in \mathbb{R}$ stands for the amount the producer pays to the data broker. Moreover, any measurable $\sigma : C \to \mathcal{S}$ is called a *segmentation scheme* (or sometimes, a *scheme*).

A mechanism (σ, τ) is incentive compatible if for all $c, c' \in C$,

$$\int_{\mathcal{D}} \pi_D(c)\sigma(\mathrm{d}D|c) - \tau(c) \geq \int_{\mathcal{D}} \pi_D(c)\sigma(\mathrm{d}D|c') - \tau(c'). \tag{IC}$$

Also, since the producer can always sell to the consumers by charging a uniform price, a mechanism (σ, τ) is individually rational if for all $c \in C$,

$$\int_{\mathcal{D}} \pi_D(c)\sigma(\mathrm{d}D|c) - \tau(c) \geq \pi_{D_0}(c). \tag{IR}$$

Henceforth, a mechanism (σ, τ) is said to be *incentive feasible* if it is incentive compatible and individually rational, and a segmentation scheme σ is said to be implementable if there exists

[8] $\overline{\boldsymbol{p}}$ is well-defined by Lemma 6 in Appendix B. Moreover, according to the notational convention stated in footnote 7, whenever $c > \max(\mathrm{supp}(D))$, $\overline{\boldsymbol{p}}_D(c) = \max \overline{V}$.

[9] For instance, if D_0 is linear demand and G is uniform; or if both D_0 and G are exponential on some intervals; or if D_0 and G are such that $D_0(v) = (1-v)^\beta$, $G(c) = c^\alpha$, for all $v \in [0,1]$, $c \in [0,1]$, for any $\alpha, \beta > 0$; or if D_0 and G take one of the aforementioned forms.

[10] Henceforth, unless otherwise noted, a *mechanism* stands for a *direct mechanism*.

a measurable $\tau : C \to \mathbb{R}$ such that (σ, τ) is incentive feasible. The goal of the data broker is to maximize expected revenue $\mathbb{E}_G[\tau(c)]$ by choosing an incentive feasible mechanism.

The data broker's revenue maximization problem exhibits several noticeable features. First, the object being allocated is infinite dimensional. After all, the data broker sells market segmentations to the producer as opposed to a one-dimensional quality or quantity variable in classical mechanism design problems (e.g., Mussa and Rosen (1978), Myerson (1981) and Maskin and Riley (1984)). In particular, it is not clear whether there exists a partial order on the space of market segmentations that would lead to the single-crossing property commonly assumed in low-dimensional screening problems. Second, the producer's outside option is type-dependent. This is because the producer has access to the consumers, and is only buying the additional information about the consumers' values.

As another remark, the model introduced above is equivalent to a model where there is one producer with private cost c and one consumer with private value v, where c and v are independently drawn from G and m^0, respectively. With this interpretation, a segmentation $s \in \mathcal{S}$ is then equivalent to a Blackwell experiment that provides the producer with information regarding the consumer's private value. Throughout the paper, the analyses and results are stated in terms of the version with a continuum of consumers, yet every statement and interpretation has an equivalent counterpart in the version with one consumer who has a private value.

4 Optimal Segmentation Design

In what follows, I characterize the data broker's optimal mechanisms. To this end, I first introduce a crucial class of mechanisms. Then I characterize the optimal mechanisms by this class.

4.1 Quasi-Perfect Segmentations and Quasi-Perfect Price Discrimination

As illustrated in the motivating example, to elicit private information from the producer, the data broker may sometimes wish to discourage sales even when there are gains from trade. In addition, the data broker would wish to extract more surplus by providing market segmentations under which all the purchasing consumers pay their values. These two features jointly lead to a specific form of market segmentation, which will be referred as *quasi-perfect segmentations*.

Definition 1. For any $c \in C$ and any $\kappa \geq c$, a segmentation $s \in \mathcal{S}$ is a *κ-quasi-perfect segmentation* for c if for s-almost all $D \in \mathcal{D}$, either $D(c) = 0$, or the set $\{v \in \text{supp}(D) : v \geq \kappa\}$ is a singleton and is a subset of $\boldsymbol{P}_D(c)$.

A κ-quasi-perfect segmentation for c is a segmentation that separates all the consumers with $v \geq \kappa$ while pooling the rest of the consumers together with them so that when the producer's marginal cost is c, every market segment with positive trading volume[11] must contain one and only one consumer-value $v \geq \kappa$ and this v is an optimal price for the producer. Notice that a κ-quasi-perfect segmentation for c induces *κ-quasi-perfect price discrimination* when the producer's marginal cost is c and when she charges the largest optimal price in (almost) all segments. Namely, a consumer with value v buys the product if and only if $v \geq \kappa$ and all purchasing consumers pay exactly their values. For instance, in the example given by Section 2, the residential data creates a 2-quasi-perfect segmentation for both $c \in \{1/4, 3/4\}$. With Definition 1, I now define the following:

Definition 2. Given any function $\psi : C \to \mathbb{R}$ with $c \leq \psi(c)$ for all $c \in C$:

1. A segmentation scheme σ is a *ψ-quasi-perfect scheme* if for G-almost all $c \in C$, $\sigma(c)$ is a $\psi(c)$-quasi-perfect segmentation for c.
2. A mechanism (σ, τ) is a *ψ-quasi-perfect mechanism* if σ is a ψ-quasi-perfect scheme and if the producer with marginal cost $\bar{c}$, when reporting truthfully, has net profit $\pi_{D_0}(\bar{c})$.

4.2 Characterization of the Optimal Mechanisms

With the definitions above, the main characterization of this paper can be stated. To this end, for any $c \in C$, define $\overline{\varphi}_G(c) := \min\{\varphi_G(c), \overline{\boldsymbol{p}}_0(c)\}$, where φ_G is the ironed virtual cost function.[12]

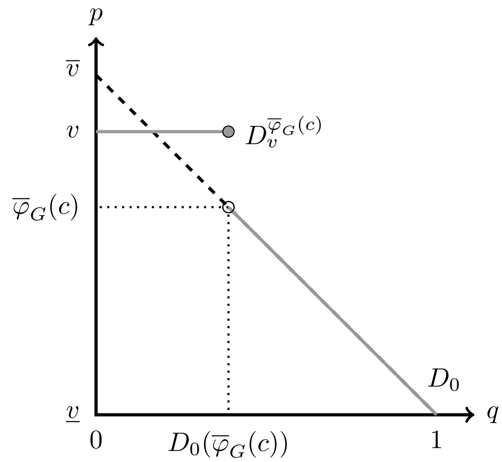
Theorem 1 (Optimal Mechanism)**.** *The set of optimal mechanisms is nonempty and is exactly the set of incentive feasible $\overline{\varphi}_G$-quasi-perfect mechanisms. Furthermore, every optimal mechanism induces $\overline{\varphi}_G(c)$-quasi-perfect price discrimination for G-almost all $c \in C$.*

From the definition of quasi-perfect segmentations, there are some degrees of freedom regarding the ways to pool the low-value consumers with the high-values. Therefore, Theorem 1 implies that there might be multiple optimal mechanisms—as long as the low-value consumers are pooled with the high-values in a way such that the mechanism is incentive feasible and is $\overline{\varphi}_G$-quasi-perfect. Nevertheless, the outcome induced by any optimal mechanism is unique. That is, under any optimal mechanism, for (almost) all marginal cost $c \in C$, a consumer with value v buys the product if and only if $v \geq \overline{\varphi}_G(c)$ and all the purchasing consumers pay their values. In other words, the multiplicity only accounts for the off-path incentives. Furthermore, there is always an explicit construction of an optimal mechanism

[11] Notice that when the producer's marginal cost is c, no trade occurs in market segment D if and only if $D(c) = 1$.

[12] Ironing in the sense of Myerson (1981).

Figure 3: Market segment $D_v^{\overline{\varphi}_G(c)}$



(see details in Appendix D). In fact, when the market demand D_0 is regular, this construction takes a particularly simple form: The low-value consumers are pooled with the high values in a way that preserves the likelihood ratios among values below the cutoff. More specifically, for any $c \in C$ and for any $v \geq \overline{\varphi}(c)$, define market segment $D_v^{\overline{\varphi}_G(c)} \in \mathcal{D}$ as

$$D_v^{\overline{\varphi}_G(c)}(p) := \begin{cases} D_0(p), & \text{if } p \in [\underline{v}, \overline{\varphi}_G(c)] \\ D_0(\overline{\varphi}_G(c)), & \text{if } p \in (\overline{\varphi}_G(c), v] \\ 0, & \text{if } p \in (v, \overline{v}] \end{cases}, \tag{2}$$

for all $p \in V$. Moreover, for any $c \in C$ and for any $p \in [\overline{\varphi}_G(c), \overline{v}]$, let

$$\sigma^*\left(\left\{D_v^{\overline{\varphi}_G(c)} : v \geq p\right\} \middle| c\right) := \frac{D_0(p)}{D_0(\overline{\varphi}_G(c))}. \tag{3}$$

In other words, for any $c \in C$, $\sigma^*(c)$ only assigns positive measure to market segments $\{D_v^{\overline{\varphi}_G(c)}\}_{v \in [\overline{\varphi}_G(c), \overline{c}]}$ and its distribution is exactly the distribution of consumers' values conditional on being above the cutoff $\overline{\varphi}_G(c)$ given by the market demand. Figure 3 illustrates σ^* by plotting the (inverse) demand[13] of a generic market segment $D_v^{\overline{\varphi}_G(c)}$ induced by $\sigma^*(c)$ (the dashed line represents the market demand D_0). This inverse demand has a jump at $D_0(\overline{\varphi}_G(c))$. To the left of $D_0(\overline{\varphi}_G(c))$, all the consumer values are concentrated at v, whereas the distribution of the consumer values to the right of $D_0(\overline{\varphi}_G(c))$ remains the same as that under D_0.

With this definition, it turns out that when D_0 is regular, as it will be shown below, there exists a unique transfer scheme $\tau^* : C \to \mathbb{R}$ such that (σ^*, τ^*) is an incentive feasible $\overline{\varphi}_G$-quasi-perfect mechanism. Thus, by Theorem 1, (σ^*, τ^*) is optimal. Henceforth, whenever D_0 is regular, I refer the mechanism (σ^*, τ^*) as the *canonical* $\overline{\varphi}_G$-quasi-perfect mechanism.

[13]See Appendix A for the formal definition of inverse demands.

Theorem 2. *Suppose that D_0 is regular. Then the canonical $\overline{\varphi}_G$-quasi-perfect mechanism (σ^*, τ^*) is optimal.*

In what follows, I will describe the main steps of the proof of Theorem 1 (which also lead to the proof of Theorem 2). Details of the proof can be found in Appendix D. Specifically, I first derive a revenue-equivalence formula and characterize the incentive compatible mechanisms. Next, I identify an upper bound $\bar{R}$ for the data broker's revenue. Then, I construct a feasible mechanism that attains $\bar{R}$, which would in turn imply every incentive feasible $\overline{\varphi}_G$-quasi-perfect mechanism is optimal. Finally, I argue that any mechanism that gives revenue $\bar{R}$ must be a $\overline{\varphi}_G$-quasi-perfect mechanism.

To highlight the main insights and avoid unnecessary complications, in this subsection, I impose some further assumptions in addition to Assumption 1. More precisely, throughout the remaining part of Section 4.2, I assume that D_0 and G are regular and that

$$\phi_G(c) \leq \overline{\boldsymbol{p}}_0(c), \, \forall c \in C. \tag{4}$$

Notice that (4) is a sufficient condition for Assumption 1. With these additional conditions, $\overline{\varphi}_G(c) = \phi_G(c)$ for all $c \in C$ and hence $\overline{\varphi}_G$ can be replaced by the virtual cost function ϕ_G. Among these assumptions, regularity of G is purely for conciseness, it can be relaxed by ironing ϕ_G. Regularity of D_0 simplifies the construction of the mechanism that attains $\bar{R}$. Without the regularity of D_0, the construction would be more involved and can be found in the Appendix D. Lastly, (4) allows a straightforward construction of the revenue upper bound $\bar{R}$. Without (4), the upper bound $\bar{R}$ may not be attainable and a tighter upper bound would be needed, which will be discussed in Section 5. Also, it is noteworthy that all the lemmas stated in this section do not rely on any of these assumptions, nor on Assumption 1.

The Revenue Equivalence Formula and an Upper Bound for Revenue

Even though the data broker's problem is more convoluted comparing to a standard monopolistic screening problem due to the high-dimensionality nature of market segmentations, a revenue-equivalence formula can still be derived by properly invoking the envelope theorem. To see this, notice that for any incentive compatible mechanism (σ, τ), the indirect utility of a producer with marginal cost c is

$$\begin{aligned} U(c) :=& \int_{\mathcal{D}} \pi_D(c)\sigma(\mathrm{d}D|c) - \tau(c) \\ =& \max_{c' \in C} \int_{\mathcal{D}} \pi_D(c)\sigma(\mathrm{d}D|c') - \tau(c') \end{aligned}$$

By the envelope theorem, the derivative of U is simply the partial derivative of the objective function evaluated at the optimum, that is,

$$U'(c) = \int_{\mathcal{D}} \pi'_D(c)\sigma(\mathrm{d}D|c).$$

Moreover, since $\pi_D(c)$ is the optimal profit of the producer with marginal cost c under segment D, again by the envelope theorem, for all $c \in C$,

$$\pi'_D(c) = -D(\overline{\boldsymbol{p}}_D(c)). \tag{5}$$

Together,

$$U(c) = U(\overline{c}) + \int_c^{\overline{c}} \left(\int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(z))\sigma(\mathrm{d}D|z) \right) \mathrm{d}z, \, \forall c \in C.$$

Therefore, under any incentive compatible mechanism (σ, τ), if a producer with marginal cost c misreports a marginal cost c' and sets prices optimally, the deviation gain can be written as

$$\begin{aligned}
& U(c) - \left(\int_{\mathcal{D}} \pi_D(c)\sigma(\mathrm{d}D|c') - \tau(c') \right) \\
= & \int_{\mathcal{D}} [\pi_D(c) - \pi_D(c')]\sigma(\mathrm{d}D|c') - (U(c) - U(c')) \\
= & \int_c^{c'} \left[\int_{\mathcal{D}} -\pi'_D(z)\sigma(\mathrm{d}D|c') - \int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(z))\sigma(\mathrm{d}D|z) \right] \mathrm{d}z \\
= & \int_c^{c'} \left[\int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(z))\sigma(\mathrm{d}D|c') - \int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(z))\sigma(\mathrm{d}D|z) \right] \mathrm{d}z
\end{aligned}$$

Together, these lead to Lemma 1 below.

Lemma 1. *A mechanism (σ, τ) is incentive compatible if and only if*

1. *For all $c \in C$,*

$$\tau(c) = \int_{\mathcal{D}} \pi_D(c)\sigma(\mathrm{d}D|c) - \int_c^{\overline{c}} \left(\int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(z))\sigma(\mathrm{d}D|z)\, \mathrm{d}z \right) - U(\overline{c}).$$

2. *For all $c, c' \in C$,*

$$\int_c^{c'} \left(\int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(z))(\sigma(\mathrm{d}D|z) - \sigma(\mathrm{d}D|c')) \right) \mathrm{d}z \geq 0.$$

Furthermore, $\overline{\boldsymbol{p}}$ can be replaced by any $\boldsymbol{p} \in \boldsymbol{P}$ for the "only if" part.

The proof of Lemma 1 can be found in Appendix D. It formalizes the heuristic arguments above by using the envelope theorem of Milgrom and Segal (2002). In essence, condition 1 in Lemma 1 is a generalized revenue-equivalence formula stating that the transfer τ must be determined by σ up to a constant, whereas condition 2 in Lemma 1 is reminiscent of Lemma 1 of Pavan et al. (2014), and relates to the integral monotonicity condition that arises in various mechanism design problems with multi-dimensional allocation spaces (see, for instance, Rochet (1987), Carbajal and Ely (2013), Pavan et al. (2014)).

From Lemma 1, for any incentive compatible mechanism (σ, τ), the data broker's expected revenue can be written as

$$\mathbb{E}_G[\tau(c)] = \int_C \left(\int_{\mathcal{D}} (\overline{\boldsymbol{p}}_D(c) - \phi_G(c))\, D(\overline{\boldsymbol{p}}_D(c))\sigma(\mathrm{d}D|c) \right) G(\mathrm{d}c) - U(\overline{c}), \tag{6}$$

which can be interpreted as the expected *virtual profit* net of a constant. That is, maximizing the data broker's expected revenue by choosing an incentive feasible mechanism (σ, τ) is equivalent to maximizing the expected virtual profit—the profit of the producer if her marginal cost c is replaced by the virtual marginal cost $\phi_G(c)$ but she still prices optimally according to marginal cost c—by choosing an implementable scheme σ.

With (6), there is an immediate upper bound for the data broker's revenue. To see this, first notice that since the producer's outside option is $\pi_{D_0}(c)$ when her cost is c, for an incentive compatible mechanism (σ, τ) to be individually rational, it must be that $U(\overline{c}) \geq \bar{\pi} := \pi_{D_0}(\overline{c})$. Moreover, notice that for any $c \in C$,

$$\begin{aligned}\int_{\mathcal{D}} (\overline{\boldsymbol{p}}_D(c) - \phi_G(c)) D(\overline{\boldsymbol{p}}_D(c))\sigma(\mathrm{d}D|c) &\leq \int_{\mathcal{D}} \max_{p\in\mathbb{R}_+}[(p - \phi_G(c))D(p)]\sigma(\mathrm{d}D|c) \\ &\leq \int_{\{v\geq\phi_G(c)\}} (v - \phi_G(c))D_0(\mathrm{d}v),\end{aligned}$$

where the second inequality holds because the last term is the total gains from trade in the economy when the producer's marginal cost is $\phi_G(c)$. Together with (6), it then follows that

$$\begin{aligned}\bar{R} :=& \int_C \left(\int_{\{v\geq\phi_G(c)\}} (v - \phi_G(c))D_0(\mathrm{d}v) \right) G(\mathrm{d}c) - \bar{\pi} \\ \geq& \int_C \left(\int_{\mathcal{D}} (\overline{\boldsymbol{p}}_D(c) - \phi_G(c))D(\overline{\boldsymbol{p}}_D(c))\sigma(\mathrm{d}D|c) \right) G(\mathrm{d}c) - U(\overline{c}) \\ =& \mathbb{E}_G[\tau(c)].\end{aligned}$$

In other words, the upper bound $\bar{R}$ is constructed by ignoring the individual rationality constrains, the global incentive compatibility constraints (i.e., condition 2 in Lemma 1) and by compelling the producer to charge prices that are optimal when her marginal cost is replaced by the virtual marginal cost.

Attaining $\bar{R}$

By the definition of quasi-perfect segmentations, for any nondecreasing function $\psi : C \to \mathbb{R}_+$ and for any ψ-quasi-perfect scheme σ, given any report $c \in C$, $\sigma(c)$ must induce $\psi(c)$-quasi-perfect price discrimination when the producer charges the largest optimal price in (almost) every segment. That is, all the consumers with $v \geq \psi(c)$ would buy the product by paying exactly their values whereas all the consumers with values $v < \psi(c)$ would not buy. As a result, all the surplus of consumers with $v \geq \psi(c)$ would be extracted and the trade volume must be the share of consumers with $v \geq \psi(c)$.[14] Specifically, for all $c \in C$,

$$\int_{\mathcal{D}} \overline{\boldsymbol{p}}_D(c) D(\overline{\boldsymbol{p}}_D(c)) \sigma(\mathrm{d}D|c) = \int_{\{v \geq \psi(c)\}} v D_0(\mathrm{d}v)$$

and

$$\int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(c)) \sigma(\mathrm{d}D|c) = D_0(\psi(c)).$$

As a result, if there is an incentive feasible ϕ_G-quasi-perfect mechanism (σ, τ), then by Lemma 1, the data broker can attain revenue

$$\begin{aligned}
\mathbb{E}[\tau(c)] &= \int_C \left(\int_{\mathcal{D}} (\overline{\boldsymbol{p}}_D(c) - \phi_G(c)) D(\overline{\boldsymbol{p}}_D(c)) \sigma(\mathrm{d}D|c) \right) G(\mathrm{d}c) - \bar{\pi} \\
&= \int_C \left(\int_{\{v \geq \phi_G(c)\}} (v - \phi_G(c)) D_0(\mathrm{d}v) \right) G(\mathrm{d}c) - \bar{\pi} \qquad (7) \\
&= \bar{R}.
\end{aligned}$$

However, not every quasi-perfect scheme is implementable. To ensure incentive compatibility, the integral inequality given by condition 2 in Lemma 1 must be satisfied. While this condition involves a continuum of constraints and is difficult to check, the following lemma provides a simpler sufficient condition.

Lemma 2. *For any nondecreasing function $\psi : C \to \mathbb{R}_+$ with $\psi(c) \geq c$ for all $c \in C$, and for any ψ-quasi-perfect scheme σ, there exists a transfer scheme $\tau : C \to \mathbb{R}$ such that (σ, τ) is incentive compatible if for any $c \in C$,*

$$\psi(z) \leq \overline{\boldsymbol{p}}_D(z), \qquad (8)$$

for (Lebesgue)-almost all $z \in [\underline{c}, c]$ and for all $D \in \mathrm{supp}(\sigma(c))$.

In essence, Lemma 2 reduces the integral inequalities given by condition 2 of Lemma 1 to pointwise conditions. Details about the proof can be found in Appendix D. The crucial step is to notice that for a ψ-quasi-perfect scheme, there are always no downward-deviation incentives (i.e., a producer with cost c would never have an incentive to misreport $c' < c$), as a higher-cost producer would find the gains from reducing the cutoff less beneficial than the increment in transfer. With this observation, as the pointwise condition (8) is sufficient to rule out upward-deviation incentives, Lemma 2 then follows.

After simplifying the incentive constraints, the following lemma then provides a crucial sufficient condition for there to exist an incentive compatible ψ-quasi-perfect mechanism.

[14] The formal arguments can be found in the proof of Lemma 11 in Appendix C.

Lemma 3. *For any nondecreasing function $\psi : C \to \mathbb{R}_+$ such that that $c \leq \psi(c) \leq \overline{\boldsymbol{p}}_0(c)$ for all $c \in C$, there exists a ψ-quasi-perfect scheme σ^* that satisfies* (8).

A direct consequence of Lemma 2 and Lemma 3 is that there exists an incentive compatible ϕ_G-quasi-perfect mechanism (σ^*, τ^*), provided that G is regular and (4) holds. Furthermore, for any $c \in C$, (4) also implies that

$$\int_c^{\overline{c}} D_0(\phi_G(z))\,\mathrm{d}z \geq \int_c^{\overline{c}} D_0(\overline{\boldsymbol{p}}_0(z))\,\mathrm{d}z.$$

Together, by Lemma 1 and (5), after possibly adding a constant to τ^* so that the indirect utility of the producer with cost $\overline{c}$ equals to $\bar{\pi}$, (σ^*, τ^*) is an incentive feasible ϕ_G-quasi-perfect mechanism, which in turn implies that (σ^*, τ^*) is optimal. Together with (7), it then follows that any incentive feasible ϕ_G-quasi-perfect mechanism is optimal.

The proof of Lemma 3 is by construction and the details can be found in Appendix D. This construction is especially simple when D_0 is regular (equivalently, when the profit function $p \mapsto (p - c)D_0(p)$ is single-peaked on $\mathrm{supp}(D_0)$ for all $c \in C$). Specifically, for any $c \in C$ and for any $v \in [\psi(c), \overline{v}]$, let $D_v^{\psi(c)} \in \mathcal{D}$ be defined as (2) with $\overline{\varphi}_G(c)$ being replaced by $\psi(c)$. Also, let $\sigma^* : C \to \Delta(\mathcal{D})$ be defined as (3) with $\overline{\varphi}_G$ being replaced by ψ. By construction, $\sigma^*(c) \in \mathcal{S}$ for all $c \in C$. Furthermore, σ^* is a ψ-quasi-perfect scheme satisfying (8). To see this, consider any $c \in C$. By regularity of D_0 and by the hypothesis that $\psi(c) \leq \overline{\boldsymbol{p}}_0(c)$, when the producer's marginal cost is c, she would prefer charging price $\psi(c)$ (or the lowest price in $\mathrm{supp}(D_0)$ that is above $\psi(c)$, if $\psi(c) \notin \mathrm{supp}(D_0)$) than charging any price $p < \psi(c)$ under D_0. Therefore, for any $v \geq \psi(c)$ and for any $p < \psi(c)$, since $D_v^{\psi(c)}(p) = D_0(p)$ and $D_v^{\psi(c)}(v) = D_0(\psi(c))$, charging price v in segment $D_v^{\psi(c)}$ must be optimal for the producer as $v \geq \psi(c)$. On the other hand, when the producer has marginal cost $z < c$, for any $v \geq \psi(c)$, since $\overline{\boldsymbol{p}}_{D_v^{\psi(c)}}$ is nonincreasing, it must be that either $\overline{\boldsymbol{p}}_{D_v^{\psi(c)}}(z) = v$ or $\overline{\boldsymbol{p}}_{D_v^{\psi(c)}}(z) < \psi(c)$. In the former case, since ψ is nondecreasing, it then follows that $\overline{\boldsymbol{p}}_{D_v^{\psi(c)}}(z) \geq \psi(c) \geq \psi(z)$. In the latter case, as $D_v^{\psi(c)}(p) = D_0(p)$ for all $p < \psi(c)$, $\overline{\boldsymbol{p}}_{D_v^{\psi(c)}}(z)$ must have been optimal for the producer under D_0 as well. Therefore, it must be that $\overline{\boldsymbol{p}}_0(z) \leq \overline{\boldsymbol{p}}_{D_v^{\psi(c)}}(z)$. Combining with the hypothesis that $\psi(z) \leq \overline{\boldsymbol{p}}_0(z)$, this then implies that $\psi(z) \leq \overline{\boldsymbol{p}}_{D_v^{\psi(c)}}(z)$. As a result, σ^* is indeed a ψ-quasi-perfect scheme satisfying (8). Notice that this also proves Theorem 2, even without the additional assumption that G is regular and (4), since $c \leq \overline{\varphi}_G(c) \leq \boldsymbol{p}_0(c)$ for all $c \in C$.

In general, for any arbitrary $D_0 \in \mathcal{D}$, the construction is more convoluted. In brief, the segmentation scheme σ^* is constructed by approximating D_0 with a sequence of step functions $\{D_n\} \subseteq \mathcal{D}$ that converges to D_0, followed by finding a desired segmentation scheme σ_n of each D_n. Together with several continuity lemmas in Appendix B, the limit of $\{\sigma_n\}$ converges to the desired segmentation scheme σ^*.

Uniqueness

To see why any optimal mechanism of the data broker is a ϕ_G-quasi-perfect mechanism, suppose that (σ, τ) is optimal. Then,

$$\begin{aligned}\bar{R} &= \int_C \left(\int_{\{v \geq \phi_G(c)\}} (v - \phi_G(c)) D_0(\mathrm{d}v) \right) G(\mathrm{d}c) - \bar{\pi} \\ &= \int_C \left(\int_{\mathcal{D}} (\overline{\boldsymbol{p}}_D(c) - \phi_G(c)) D(\overline{\boldsymbol{p}}_D(c)) \sigma(\mathrm{d}D|c) \right) G(\mathrm{d}c) - \bar{\pi}, \end{aligned} \tag{9}$$

which in turn implies that for (almost) all $c \in C$,

$$\int_{\{v \geq \phi_G(c)\}} (v - \phi_G(c)) D_0(\mathrm{d}v) = \int_{\mathcal{D}} (\overline{\boldsymbol{p}}_D(c) - \phi_G(c)) D(\overline{\boldsymbol{p}}_D(c)) \sigma(\mathrm{d}D|c), \tag{10}$$

since the left-hand side is the efficient surplus in an economy where the producer's cost is $\phi_G(c)$ and hence must be an upper-bound of the right-hand side, and (9) implies that the right-hand side must attain this upper bound.

It then follows σ must be a ϕ_G-quasi-perfect mechanism. Indeed, if σ is not a ϕ_G-quasi-perfect scheme, it must be that there is a positive G-measure of $c \in C$ and a positive $\sigma(c)$-measure of $D \in \mathrm{supp}(\sigma(c))$ such that either $D(v) > 0$ for some $v > \overline{\boldsymbol{p}}_D(c)$, or $D(\phi_G(c)) \neq D(\overline{\boldsymbol{p}}_D(c))$. That is, either there are some consumers with $v \geq \phi_G(c)$ who do not buy the product or buy the product at a price below v, or there are some consumers with $v < \phi_G(c)$ who end up buying the product. This contradicts (10). As a result, (σ, τ) must be a ϕ_G-quasi-perfect mechanism. Moreover, (σ, τ) must also induce quasi-perfect price discrimination since $\overline{\boldsymbol{p}}$ can be replaced with any $\boldsymbol{p} \in \boldsymbol{P}$ according to Lemma 1.

4.3 Further Remarks and Implementation

Theorem 1 underlines a noteworthy feature of the optimal mechanisms. According to Theorem 1, for any optimal mechanism (σ, τ), the segmentation scheme σ does not generate value-revealing segmentations in general. Specifically, for any report c such that $\overline{\varphi}_G(c) > \underline{v}$, there are market segments $D \in \mathrm{supp}(\sigma(c))$ containing consumers with distinct values. The reason is that in order to incentivize the producer to set prices in desirable ways and to elicit information from the producer, some market segments must contain consumers with values below the desirable threshold $\overline{\varphi}_G(c)$. By pooling the high-value consumers with the low-value ones in the same market segment while separating them from other high-value consumers, the data broker is able to incentivize the producer to set prices at the highest value in each market segment and induce $\overline{\varphi}_G(c)$-quasi-perfect price discrimination for all c, which in turn enables the data broker to elicit private information by discouraging trade and extract surplus from the purchasing consumers at the same time. This also means it is not optimal for the data broker to release all the information about consumers' values.

As an example for an optimal mechanism, consider the case where D_0 is linear and G is a uniform distribution with $V = C = [0, 1]$. It then follows $\overline{\varphi}_G(c) = 2c$ for all $c \in [0, 1/3]$ and $\overline{\varphi}_G(c) = (1 + c)/2$ for all $c \in (1/3, 1]$. In this case, the canonical $\overline{\varphi}_G$-quasi-perfect mechanism is described by a uniform distribution on the market segments $\{D_v^{\overline{\varphi}_G(c)}\}_{v \in [\overline{\varphi}_G(c), 1]}$, where each market segment $D_v^{\overline{\varphi}_G(c)}$ is defined by (2).

As another example, notice that in the motivating example (Section 2), the optimal menu $\mathcal{M}^*$, which consists of the value-revealing data (with a price of 7/12) and the residential data (with a price of 1/3), implements the canonical quasi-perfect mechanism with a desirable cutoff function. Indeed, the residential data induces a 2-quasi-perfect segmentation for $c = 3/4$ as it only separates the high-value consumers (graduate students and professors) and pools the low-value consumers (undergraduate students) with them while preserving their mass. On the other hand, the value-revealing data induces a 1-quasi-perfect segmentation for $c = 1/4$. According to the characterization above, since market demand D_0 is regular and since the virtual costs are 1/4 and 5/4 (for costs 1/4 and 3/4, respectively),[15] the menu $\mathcal{M}^*$ is indeed optimal.

Furthermore, the quasi-perfect mechanisms (and thus the optimal mechanisms) can be implemented by a menu of price-recommendation data, and the implementation is particularly straightforward for the canonical $\overline{\varphi}_G$-quasi-perfect mechanism (σ^*, τ^*) when D_0 is regular. More specifically, to implement (σ^*, τ^*), the data broker can first generate an auxiliary characteristic $u \in [0, 1]$ independently according to a uniform distribution for each consumer. Then, he can create a menu in which each item is indexed by $c \in C$. Upon choosing item $c \in C$, the producer has to pay $\tau^*(c)$ to the data broker. The data broker then segments the consumers according to the partition $\{C_p\}_{p \in [\overline{\varphi}_G(c), \overline{v}]}$, as illustrated by Figure 4, where for each $p \in [\overline{\varphi}_G(c), \overline{v}]$

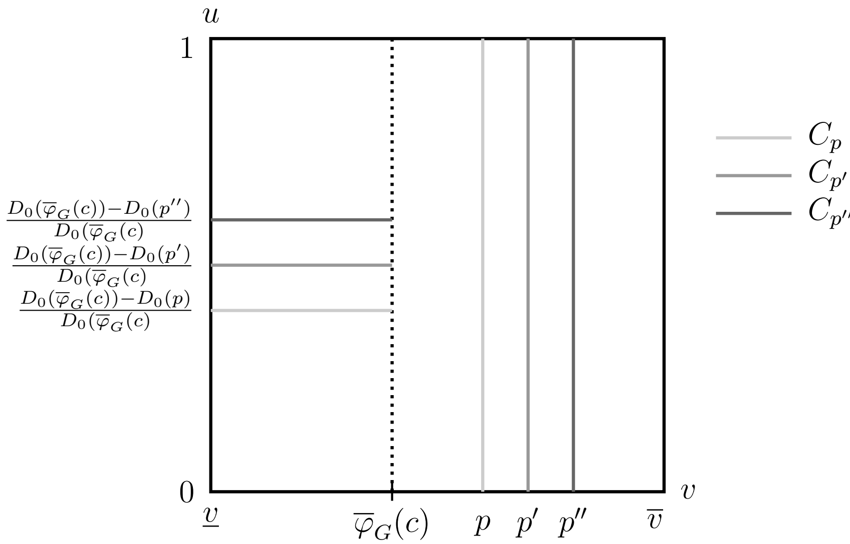
$$C_p := \{(v, u) : v = p\} \bigcup \left\{ (v, u) : v < \overline{\varphi}_G(c),\, u = \frac{D_0(\overline{\varphi}_G(c)) - D_0(p)}{D_0(\overline{\varphi}_G(c))} \right\}$$

is the set of consumers who either have value $v = p$, or have value $v < \overline{\varphi}_G(c)$ and auxiliary characteristic $u = (D_0(\overline{\varphi}_G(c)) - D_0(p))/D_0(\overline{\varphi}_G(c))$. Finally, for every $p \in [\overline{\varphi}_G(c), \overline{v}]$, the consumers belonging to C_p are attached with a recommended price p and the producer receives a list of recommended prices, one for each consumer.

It can be verified that the producer would choose item c when her marginal cost is c. Moreover, among the consumers who are attached with recommended price p, their values are distributed according to $D_p^{\overline{\varphi}_G(c)}$, which then implies it is optimal for the producer to charge them price p. As such, the mechanism (σ^*, τ^*) is implemented.

[15] Although the characterization is stated for cost distributions that admit densities, as in standard mechanism design problems, there is a straightforward analogous notion of virtual cost function when the cost distribution has atoms.

Figure 4: Price-Recommendation Data



5 Consequences of Consumer-Data Brokership

5.1 Surplus Extraction

One of the most pertinent questions about consumer-data brokership is how it affects consumer surplus. Are the data broker's possession of consumer data and the ability to sell them to a producer detrimental for the consumers? If so, to what extent? Meanwhile, can the consumers benefit from the fact that the data broker does not have access to the consumers and only affects the product market indirectly by selling data to the producer? While currently being a focus of policy debates, the following result, as an implication of Theorem 1, answers a certain aspect of this question.

Theorem 3 (Surplus Extraction)**.** *Consumer surplus is zero under any optimal mechanism.*

Theorem 3 follows directly from the characterization given by Theorem 1. Indeed, according to Theorem 1, any optimal mechanism must induce $\overline{\varphi}_G(c)$-quasi-perfect price discrimination for (almost) all $c \in C$, which means that every purchasing consumer must be paying their values. Notably, Theorem 3 provides an unambiguous and substantial assertion about the consumer surplus under data brokership. According to Theorem 3, even though the data broker does not sell the product to the consumers directly and only affects the market by creating market segmentations for the producer, it is as if the consumers are perfectly price discriminated and all the surplus is extracted away (even though the optimal mechanisms do not induce perfect price discrimination in general). This means that as long as the data broker possesses consumer data and can sell them to a producer, from the consumers' perspective, it is the same as buying the product from a monopolist who can implement perfect price discrimination. More practically, this result means it is impossible to expect the consumers to benefit from the gap between the ownership of production technology and ownership of consumer data.

5.2 Comparisons with Uniform Pricing

Although Theorem 3 indicates data brokership is undesirable for the consumers, it does not imply that data brokership is detrimental to the entire economy. After all, by facilitating price discrimination, data brokership may increase total surplus comparing to uniform pricing where no information about the consumers' values is revealed. Theorem 1, together with Proposition 1, allows such a comparison.

Proposition 1. *The data broker's optimal revenue is no less than the consumer surplus under uniform pricing.*

An immediate consequence of Proposition 1 is that total surplus under data brokership is greater compared with uniform pricing, as summarized below.

Theorem 4 (Total Surplus Improvement)**.** *Data brokership always increases total surplus compared with uniform pricing.*

Theorem 4 means that even though data brokership is extremely harmful to the consumers, in terms of total surplus it creates, however, it is always better than the environment where no information about the consumers' values can be disclosed.

Another implication of Proposition 1 pertains to the source of consumer data. So far, it has been assumed that the data broker owns all the consumer data and is able to perfectly predict each consumer's value. In contrast, a different ownership structure of consumer data can be considered. In this setting, the data broker does not have any data in the first place and has to purchase them from the consumers.[16] Proposition 1 immediately implies that, if the data broker has to purchase data by compensating the consumers with monetary transfers *before* they learn their values,[17] then the optimal mechanism would be to purchase

[16]For simplicity, a "purchase" of data here means that the data broker gains access to *all* the consumer data, in the sense that he can provide any segmentation of D_0 to the producer once he makes the purchase. In the Supplemental Material, I further extend the model and allow the data broker to make a take-it-or-leave-it offer to purchase *any* kind of consumer data and then sell them to the producer. (i.e., offer any segmentation of D_0 that is a mean-preserving contraction of the segmentation induced by the purchased data.)

[17]It is crucial here the data broker purchases *before* the consumers learn their value, since otherwise he would also have to screen the consumers to elicit their private information. Such ex-ante purchase of consumer data is plausibly suitable for online activities. After all, in online settings, consumers often do not consider their values about a particular product when they agree that their personal data such as browsing histories, IP address and cookies, can be collected by the data brokers. Nevertheless, other purchase timing would also be a relevant question, which can be explored in future research.

all the data by paying the consumers their ex-ante surplus under uniform pricing and then use any optimal mechanism characterized by Theorem 1 to sell these data to the producer. Furthermore, since the data broker's revenue is greater than the consumer surplus under uniform pricing according to Proposition 1, and since the producer always has an outside option of uniform pricing, this outcome is in fact Pareto improving compared with uniform pricing in the ex-ante sense, as stated below.[18]

Theorem 5 (Data Ownership). *If the consumers own their data and if the data broker can purchase data from the consumers before they learn their values, then data brokership is Pareto improving compared with uniform pricing in the ex-ante sense.*

5.3 Comparisons across Market Regimes

In addition to its welfare implications, the characterization of Theorem 1 provides further insights about the comparisons across different regimes of the market. Indeed, other than selling consumer data to the producer, there are several other market regimes under which the data broker can profit from the consumer data he owns. Therefore, it would be policy-relevant to compare the outcomes induced by these different market regimes. In what follows, I consider several other market regimes in addition to data brokership, including **vertical integration**, **direct acquisition**, **exclusive retail**, and **price-controlling data brokership**. Then, I compare the implications among these different regimes using the characterization provided by Theorem 1.

Vertical Integration— The producer's marginal cost of production becomes common knowledge (for exogenous reasons such as regulation or technological improvements) and the data broker vertically integrates with the producer. That is, the vertically integrated entity is able to produce the product and sell to the consumers via perfect price discrimination.

Direct Acquisition— The producer's marginal cost of production is still private, but the data broker is allowed to acquire the producer by offering her a lump-sum transfer $t \geq 0$. If the producer rejects the offer, she receives her optimal uniform pricing profit and the data broker receives zero. If the producer accepts the offer, she receives t and the data broker acquires the production technology, learns the marginal cost of production, and is able to perfectly price discriminate the consumers.

[18] Jones and Tonetti (2020) also conclude that granting consumers ownership of their own data is welfare-improving. However, their results are derived in a monopolistic competition setting and the main driving force is the non-rival property of data, whereas Theorem 5 is derived under a monopoly setting and the main rationale is that consumer data facilitate price discrimination, which in turn enhance efficiency.

Exclusive Retail— The producer's marginal cost of production remains private. The data broker negotiates with the producer to purchase the product as well as the exclusive right to sell the product. That is, the data broker can offer a menu, where each item in this menu specifies the quantity $q \in [0,1]$ that the producer has to produce and supply to the data broker, as well as the amount of payment the data broker has to pay to the producer t. If the producer chooses an item (q,t) from this menu, she receives profit $t - cq$ while the data broker pays t and can sell at most q units exclusively to the consumers through any market segmentation and at any prices. If the producer rejects this menu, she retains her optimal uniform profit and the data broker receives zero.

Price-Controlling Data Brokership— The producer's marginal cost of production is private information. The data broker, in addition to being able to create market segmentations and sell them to the producer, can further specify what price should be charged in each market segment as a part of the contract. If the producer rejects, she retains her optimal uniform pricing profit and the data broker receives zero. That is, the data broker offers a mechanism $(\sigma, \tau, \boldsymbol{\gamma})$ such that for all $c, c' \in C$,

$$\int_{\mathcal{D}\times\mathbb{R}_+} (p-c)D(p)\boldsymbol{\gamma}(\mathrm{d}p|D,c)\sigma(\mathrm{d}D|c) - \tau(c) \geq \int_{\mathcal{D}\times\mathbb{R}_+} (p-c)D(p)\boldsymbol{\gamma}(\mathrm{d}p|D,c')\sigma(\mathrm{d}D|c') - \tau(c')$$

and for all $c \in C$,

$$\int_{\mathcal{D}\times\mathbb{R}_+} (p-c)D(p)\boldsymbol{\gamma}(\mathrm{d}p|D,c)\sigma(\mathrm{d}D|c) - \tau(c) \geq \pi_{D_0}(c),$$

where for each $c \in C$, $\sigma(c) \in \mathcal{S}$ is the market segmentation provided to the producer, $\tau(c) \in \mathbb{R}$ is the payment from the producer to the data broker, and $\boldsymbol{\gamma}(c) : \mathcal{D} \to \Delta(\mathbb{R}_+)$ is a transition kernel so that $\boldsymbol{\gamma}(\cdot|D,c)$ specifies the distribution which prices charged in segment D must be drawn from.

With these definitions, for each market regime, there is an associated profit maximization problem. Henceforth, two market regimes are said to be *outcome-equivalent* if every solution of the profit maximization problems associated with either market regime induces the same market outcome (i.e., consumer surplus, producer's profit, data broker's revenue and the allocation of the product).

An immediate consequence of Theorem 1 is the comparison between data brokership and vertical integration. To see this, recall that any optimal mechanism (σ, τ) of the data broker must induce $\overline{\varphi}_G$-quasi-perfect price discrimination but not perfect price discrimination in general, as $\overline{\varphi}_G(c) > c$ for all $c > \underline{c}$. Thus, whenever there are some consumers with values between c and $\overline{\varphi}_G(c)$ for a positive measure of c, any optimal mechanism would not lead to an efficient allocation because there would be some consumers who end up not buying the product even though their values are greater than the marginal cost. Together with Theorem 3, this means that vertical integration between the data broker and producer increases total surplus while leaving the consumer surplus unchanged when D_0 has full support on V and there is no common knowledge of gains from trade. After all, consumer surplus is always zero under both regimes, whereas the integrated entity after vertical integration does not create any friction and would perfectly price-discriminate the consumers whose values are above the marginal cost.

Theorem 6 (Vertical Integration)**.** *Compared with data brokership, vertical integration increases total surplus and leaves the consumer surplus unchanged if D_0 is strictly decreasing and $\underline{v} < \bar{c}$.*

To compare other market regimes, it is noteworthy that since prices are contractable under price-controlling data brokership, for any mechanism $(\sigma, \tau, \boldsymbol{\gamma})$, the producer's private marginal cost affects her profit only through the quantity produced and sold to the consumers induced by $(\sigma, \boldsymbol{\gamma})$. This effectively reduces allocation space under price-controlling data brokership to a one-dimensional quantity space, which is the same as the allocation space under exclusive retail. In fact, as stated in Lemma 4 below, price-controlling data brokership is always equivalent to exclusive retail.

Lemma 4. *Exclusive retail and price-controlling data brokership are outcome-equivalent.*

With Lemma 4, to compare exclusive retail and price-controlling data brokership with data brokership, it suffices to compare only price-controlling data brokership with data brokership. This comparison is particularly convenient since the price-controlling data broker's revenue maximization problem is a relaxation of the data broker's. After all, with the extra ability to contract on prices, the constraints in the price-controlling data broker's problem must be weaker. Nevertheless, as an implication of Theorem 1 and Proposition 2 below, it turns out that the data broker's optimal revenue is in fact the same as the price-controlling data broker's optimal revenue.

Proposition 2. *Any optimal mechanism of the price-controlling data broker induces $\bar{\varphi}_G(c)$-quasi-perfect price discrimination for G-almost all $c \in C$. In particular, the optimal revenue is*

$$R^* = \int_C \left(\int_{\{v \geq \overline{\varphi}_G(c)\}} (v - \phi_G(c)) D_0(\mathrm{d}v) \right) G(\mathrm{d}c) - \bar{\pi}.$$

According to Theorem 1 and Lemma 1, the optimal revenue of the data broker must also be R^*. This means that the additional ability to control prices does not benefit the data broker at all. In fact, as stated by Theorem 7 below, this ability is entirely irrelevant in terms of market outcomes.

Theorem 7 (Outcome-Equivalence)**.** *Exclusive retail, price-controlling data brokership and data brokership are outcome-equivalent.*

In other words, Theorem 7 means that even though the data broker only affects the product market indirectly by selling consumer data, the market outcomes he induces are the same as those when he has more control over the product market by either becoming a price-controlling data broker or an exclusive retailer. More specifically, from the data broker's perspective, having control over how the product is sold in addition to consumer data adds no extra values to his revenue. As for the producer, her profit in face of a data broker is the same as if she sells the product, as well as the exclusive right to sell the product, to this data broker. Preserving the access to consumers and the right to sell the product is in fact not more profitable. In addition, the allocation of the product induced by a data broker is the same as that induced by an exclusive retailer. Therefore, the channel through which the product is sold to the consumers does not affect the amount of products being produced, nor does it affect to whom the product is sold.

This outcome-equivalence result has several further implications. First, it implies that there are no incentives for the data broker to become more active, as the data broker's revenue would remain the same even if he becomes a price-controlling data broker or an exclusive retailer. Second, from a policymaker's perspective, it means that no further concerns should be raised even if a data broker eventually becomes more active. After all, the market outcomes and the amount of deadweight loss would remain the same.

As another remark, the fact that the price-controlling data broker's optimal revenue R^* is an upper bound for the data broker's optimal revenue completes the intuition behind the proof of Theorem 1 without the additional assumption (4) imposed in Section 4.2. To see this, since the price-controlling data broker's optimal mechanisms always induce $\overline{\varphi}_G$-quasi-perfect price discrimination for (almost) all $c \in C$ according to Proposition 2, proving Theorem 1 is essentially reduced to finding an incentive feasible $\overline{\varphi}_G$-quasi-perfect mechanism. Furthermore, by the definition of $\overline{\varphi}_G$, $c \leq \overline{\varphi}_G(c) \leq \overline{\boldsymbol{p}}_0(c)$ for all $c \in C$, and hence $\overline{\varphi}_G$ satisfies the condition required by Lemma 3. As a result, combining Lemma 2 and Lemma 3, there is indeed an incentive feasible $\overline{\varphi}_G$-quasi-perfect mechanism.

Finally, to compare direct acquisition with other market regimes, first notice that under direct acquisition, given a proposed transfer t, the producer would accept this offer if and only if $\pi_{D_0}(c) \leq t$. As a result, the profit maximization problem associated with direct acquisition is

$$\max_{t \geq 0} \int_{\{c \in C: \pi_{D_0}(c) \leq t\}} \left(\int_{\{v \geq c\}} (v - c) D_0(\mathrm{d}v) - t \right) G(\mathrm{d}c).$$

By Theorem 7, it suffices to compare direct acquisition with exclusive retail. Compared with exclusive retail, on the one hand, direct acquisition allows the data broker to learn the marginal cost of production conditional on acquiring the producer, whereas the exclusive retailer can never completely learn the marginal cost but can only screen the producer to elicit this private information. On the other hand, direct acquisition creates severe adverse selection as only when the publisher's marginal cost is high enough would she be willing to sell the production technology at a given price. As a result, even with successful acquisition, the data broker can only operate with a relatively inferior production technology. Consequently, whether direct acquisition is more profitable than exclusive retail depends on the distribution of marginal cost G and the market demand D_0. Nevertheless, with the complete characterization given by Theorem 1, direct acquisition can be compared with data brokership through direct calculations.

Example 1. Suppose that $\mathrm{supp}(D_0) = \mathrm{supp}(G) = [0,1]$ and $D_0(p) = (1-p)$, $G(c) = c$ for all $p, c \in [0,1]$. Then $\phi_G(c) = 2c$ for all $c \in C$, $\overline{\varphi}_G(c) = 2c$ for all $c \in [0,1/3]$ and $\overline{\varphi}_G(c) = (1+c)/2$ for all $c \in (1/3,1]$. Also, $\pi_{D_0}(c) = (1-c)^2/4$. In this case,

$$R^* = \int_0^{\frac{1}{3}} \left(\int_{2c}^{1} (v-2c)\,\mathrm{d}v \right) \mathrm{d}c + \int_{\frac{1}{3}}^{1} \left(\int_{\frac{1+c}{2}}^{1} (v-2c)\,\mathrm{d}v \right) \mathrm{d}c = \frac{1}{18}.$$

In contrast, for any $t \geq 0$,

$$\int_{\{c \in C: \frac{(1-c)^2}{4} \leq t\}} \left(\int_{\{v \geq c\}} (v-c)\,\mathrm{d}v - t \right) \mathrm{d}c \leq 0.$$

As a result, data brokership (and hence price-controlling data brokership and exclusive retail) yields more revenue to the data broker than direct acquisition.

Example 2. Suppose that $\mathrm{supp}(D_0) = \mathrm{supp}(G) = [0,1]$ and $D_0(p) = (1-p)$, $G(c) = c^{\frac{1}{n}}$ for all $p, c \in [0,1]$, for some $n \in \mathbb{N}$. Then $\phi_G(c) = (n+1)c$ for all $c \in C$, $\overline{\varphi}_G(c) = (n+1)c$ for all $c \in [0, 1/(1+2n)]$ and $\overline{\varphi}_G(c) = (1+c)/2$ for all $c \in (1/(1+2n), 1]$. Also, $\pi_{D_0}(c) = (1-c)^2/4$. In this case, for n large enough ($n > 15$ to be more precise),

$$R^* < \int_0^1 \left(\int_{\{v \geq c\}} (v-c)\,\mathrm{d}v - \frac{1}{4} \right) \frac{1}{n} c^{-\frac{n-1}{n}}\,\mathrm{d}c.$$

Hence, for n large enough, data brokership (and hence price-controlling data brokership and exclusive retail) yields less revenue to the data broker than direct acquisition.

6 Extensions

6.1 Sufficient Conditions and Relaxations of Assumption 1

Despite being a technical condition, Assumption 1 has an economically interpretable sufficient condition (4). To better understand this, recall that by definition, $\phi_G(c) = c + G(c)/g(c)$, and therefore $\phi_G(c)$ is the actual marginal cost c plus the information rent $G(c)/g(c)$. On the other hand, instead of regarding $\overline{\boldsymbol{p}}_0(c)$ as the optimal uniform price for the producer when her marginal cost is c, $\overline{\boldsymbol{p}}_0(c)$ can be written as $\overline{\boldsymbol{p}}_0(c) = c + \xi_0(c)$, where $\xi_0(c) := \overline{\boldsymbol{p}}_0(c) - c$ is the *monopoly mark-up* that the producer charges under uniform pricing. From this perspective, (4) is equivalent to

$$\frac{G(c)}{g(c)} \leq \xi_0(c),\ \forall c \in C.$$

That is, the information rent that the producer retains due to asymmetric information about her marginal cost is less than her monopoly mark-up.

Furthermore, (4) can also be interpreted as the gains from trade being large enough. More specifically, for any demand $D_0 \in \mathcal{D}$, define a location family $\{D_0^k\}_{k \geq 0}$ by moving the support of D_0 without changing the shape of the distribution. That is, $D_0^k(p) := D_0(p-k)$ for all $p \in V$ and for all $k > 0$. Within this family, it is natural to rank the gains from trade by the location parameter k. In the Supplemental Material, I show that there exists $\overline{k} \geq 0$ such that (4) holds if and only if $k \geq \overline{k}$.[19]

Although the results introduced above rely on Assumption 1, the sole purpose of Assumption 1 is to ensure that as a revenue upper bound, the price-controlling data broker's problem has a closed form solution. After all, by Lemma 4, the price-controlling data broker's problem is essentially a nonlinear screening problem with one-dimensional allocation space and type-dependent outside options. A common feature of such problems is that the characterization of the optimal mechanisms involves Lagrange multipliers in general (see, for instance, Lewis and Sappington (1989) and Jullien (2000)). Assumption 1, however, yields a closed form solution for the price-controlling data broker's problem (Proposition 2), which in turn allows an explicit construction of an incentive feasible mechanism for the data broker that attains the revenue upper bound.

Consequently, many of the results can be extended to environments without Assumption 1. First, Theorem 3 actually does not rely on Assumption 1 at all. A strengthened version of Theorem 3 can be found in the Supplemental Material, which ensures both the existence of an optimal mechanism for the data broker and the fact that any optimal mechanism must yield zero consumer surplus. A crucial step of the proof is to take any mechanism (σ, τ) under which the consumers retain positive surplus and apply Lemma 3 to every market segment $D \in \mathrm{supp}(\sigma(c))$ for every report c, with the cutoff function ψ being $\overline{\boldsymbol{p}}_D$. This would induce another segmentation scheme. The fact that all the market segments $D \in \mathrm{supp}(\sigma(c))$ are decomposed according to a $\overline{\boldsymbol{p}}_D(c)$-quasi-perfect segmentation and the hypothesis that consumers retain positive surplus under (σ, τ) yield a strict improvement on the data broker's

[19] Clearly, if k is large enough so that $\underline{v} \geq \phi_G(\overline{c})$, then there is common knowledge of gains from trade even after incorporating the information rents and hence the value-revealing scheme would be optimal. In the Supplemental Material, I show that there exists k such that (4) holds even then $\underline{v} < \phi_G(\overline{c})$.

revenue. Moreover, (8) ensures that such decomposition relaxes the incentive compatibility and individual rationality constraints.

In addition, the main characterization (Theorem 1) can be generalized as well. More specifically, in the Supplemental Material, I show that as long as D_0 is continuous, there exists a nondecreasing function φ^* such that every optimal mechanism must be a φ^*-quasi-perfect mechanism. However, unlike $\overline{\varphi}_G$, the cutoff function φ^* cannot be concisely defined by the model primitives. Nevertheless, the proof of this result gives a (partial) characterization of this cutoff function. A crucial property of this cutoff function φ^* is that $\varphi^*(c) > c$ for all $c > \underline{c}$. Therefore, together with the fact that consumer surplus must be zero under any optimal mechanism, the characterization of optimal mechanisms guarantees the validity of the comparison between vertical integration and data brokership. That is, as long as D_0 is continuous, Theorem 6 does not require Assumption 1 either. Finally, without Assumption 1, the outcome equivalence result Theorem 7 may not hold. The price-controlling data broker's optimal revenue is sometimes strictly greater than the data broker's. However, from the characterization of the optimal cutoff function φ^*, it can be shown that price-controlling data brokership (and hence exclusive retail) Pareto-dominates data brokership whenever D_0 is continuous.

6.2 Consumers' Private Information

Given the amount of consumer data that can be collected, their predictive power is approaching perfect estimations of consumers' values. Nonetheless, it is still imperative to explore the economic implications of the possibility when the consumers have some private information. This section extends the baseline model in Section 3 and allows the consumers to retain some pieces of information.

To formally model this, let Θ be a Polish space that denotes a set of consumer characteristics which can be disclosed by the data broker. Suppose that among the consumers, their available characteristics θ are distributed according to $\beta_0 \in \Delta(\Theta)$. These characteristics are informative about the consumers' values but there may still be variation in values even among the consumers who share the same characteristics. Specifically, given any $\theta \in \Theta$, suppose that among the consumers who share characteristic θ, their values are distributed according to $m^\theta \in \Delta(V)$ and m^θ induces a demand $D_\theta \in \mathcal{D}$ (i.e., $D_\theta(p) := m^\theta([p, \overline{v}])$ for all $p \in P$) for each $\theta \in \Theta$. The data broker is only able to segment the market according to θ but not v. In this environment, a market segmentation is then defined by $s \in \Delta(\Delta(\Theta))$ such that

$$\int_{\Delta(\Theta)} \beta(A) s(\mathrm{d}\beta) = \beta_0(A),$$

for any measurable $A \subseteq \Theta$. As a result, there is now a limit on how predictive the data can be and the environment is described by $(\{D_\theta\}_{\theta \in \Theta}, \beta_0)$.

To simplify analyses, I further specialize the environment. Suppose that there are finitely many possible characteristics. That is, $|\Theta| < \infty$. Moreover, suppose that $\{\text{supp}(D_\theta)\}_{\theta\in\Theta}$ forms a partition of V and $\text{supp}(D_\theta)$ is an interval for all $\theta \in \Theta$. This specialization will be referred as *partitional*. In other words, the data broker only has partial information about the consumers' values and can at most identify which interval a consumer's value belongs to. Even when θ is perfectly revealed, the producer would still be unable to identify each consumer's value. In this environment, the market demand D_0 is given by

$$D_0(p) := \sum_{\theta\in\Theta} D_\theta(p)\beta_0(\theta),$$

for all $p \in V$. Moreover, a market segmentation s induces market segments $\{D_\beta\}_{\beta\in\text{supp}(s)}$ and

$$\sum_{\beta\in\text{supp}(s)} D_\beta(p)s(\beta) = D_0(p),$$

for all $p \in V$, where $D_\beta(p) := \sum_{\theta\in\Theta} D_\theta(p)\beta(\theta)$ for any $\beta \in \Delta(\Theta)$ and any $p \in V$.

When the consumers' values can never be fully disclosed, it is clear that their surplus will increase. After all, it is no longer possible for the producer to charge the consumers their values as the additional variation in values given by D_θ always allows some consumers to buy the product at a price that is below their values. Nevertheless, as shown in Theorem 8, under any optimal mechanism, consumer surplus must be lower than the case when all the information about θ is revealed to the producer. That is, the main implication of Theorem 3—for the consumers, the presence of a data broker is no better than a scenario where their data is fully revealed to the producer—is still valid even when the consumers retain some private information through a partitional $(\{D_\theta\}_{\theta\in\Theta}, \beta_0)$.

Theorem 8. *For any partitional $(\{D_\theta\}_{\theta\in\Theta}, \beta_0)$ and any distribution of marginal cost G, an optimal mechanism exists. Furthermore, the consumer surplus under any optimal mechanism of the data broker is lower than the case when θ is fully disclosed.*

The intuition behind Theorem 8 is simple. Since there are only finitely many characteristics, identifying the consumers' characteristic θ effectively enables the producer to categorize the consumers into finitely many "blocks" so that every possible value belongs to one and only one block. As a result, when changing prices within each block of values, the trading volume is only affected by purchasing decisions of the consumers whose values are within that block. Such separability allows an analogous argument as in the proof of the generalized version of Theorem 3 (provided in the Supplemental Material) which shows that the data broker can always construct a mechanism that increases its revenue if the consumer surplus is higher than that when the characteristic θ is not full-revealed.

In addition to the surplus extraction result, the characterization of the optimal mechanisms can be generalized as well. That is, with proper regularity conditions, there is an optimal mechanism that is analogous to the canonical $\overline{\varphi}_G$-quasi-perfect mechanism introduced in Section 4. To state this result, given any partitional $(\{D_\theta\}_{\theta\in\Theta}, \beta_0)$, for each $\theta \in \Theta$, write $\mathrm{supp}(D_\theta)$ as $[l(\theta), u(\theta)]$. For any $p \in V$, let $\theta_p \in \Theta$ be the unique θ such that $p \in (l(\theta), u(\theta)]$. For any $c \in C$, let $\widehat{\boldsymbol{p}}_0(c)$ be the largest optimal price for the producer with marginal cost $c \in C$ under the demand whose support contains $\overline{\boldsymbol{p}}_0(c)$.[20] Also, let $\widehat{\varphi}_G(c) := \min\{\varphi_G(c), \widehat{\boldsymbol{p}}_0(c)\}$ for all $c \in C$. Furthermore, given any function $\psi : C \to \mathbb{R}_+$, say that a mechanism (σ, τ) is a canonical ψ-quasi-perfect segmentation if the producer with marginal cost $\bar{c}$, when reporting truthfully, recevies $\bar{\pi}$, and if for any $c \in C$, and for any $\beta \in \mathrm{supp}(\sigma(c))$, either

$$\beta(\theta') = \beta^{\theta}_{\psi(c)}(\theta') := \begin{cases} \beta_0(\theta'), & \text{if } u(\theta') < \psi(c) \text{ and } u(\theta) \geq \psi(c) \\ \sum_{\{\hat{\theta}: u(\hat{\theta}) \geq \psi(c)\}} \beta_0(\hat{\theta}), & \text{if } u(\theta') \geq \psi(c) \text{ and } \theta' = \theta \\ 0, & \text{otherwise} \end{cases}, \tag{11}$$

for any $\theta, \theta' \in \Theta$, or

$$\mathrm{supp}(\beta) = \{\theta' : l(\theta') \leq \psi(c)\} \cup \{\theta\} \tag{12}$$

for some $\theta \in \Theta$ with $l(\theta) \geq \psi(c)$ and

$$\beta(\theta') = \beta_0(\theta'). \tag{13}$$

for all $\theta' \in \Theta$ such that $u(\theta') < \psi(c)$.

With these definitions, Theorem 9 below prescribes an optimal mechanism for the data broker.

Theorem 9. *For any partitional $(\{D_\theta\}_{\theta\in\Theta}, \beta_0)$ and any distribution of marginal cost G such that the function $c \mapsto \max\{(\phi_G(c) - \widehat{\boldsymbol{p}}_0(c)), 0\}$ is nondecreasing and that D_0 is regular, there is a canonical $\widehat{\varphi}_G$-quasi-perfect mechanism that is optimal.*

6.3 Targeted Marketing

So far, the discussions have been abstracting away the possibility that the data broker can use consumer data to facilitate targeted marketing by assuming there is only one product. In fact, one of the most common arguments in favor of the usage and provision of consumer data is that it also benefits the consumers because more relevant products can be advertised to the consumers and therefore more surplus can be created. The following extension explores this aspect.

[20] That is, $\widehat{\boldsymbol{p}}_0(c) := \overline{\boldsymbol{p}}_{D_{\theta_{\overline{\boldsymbol{p}}_0(c)}}}(c)$. Notice that $\widehat{\boldsymbol{p}}_0(c) \leq \overline{\boldsymbol{p}}_0(c)$ for all $c \in C$. Moreover, in the case where the data broker can disclose all the information about the value v, $\widehat{\boldsymbol{p}}_0(c) = \overline{\boldsymbol{p}}_0(c)$ for all $c \in C$.

Formally, suppose that, instead of a single product, there are $J \in \mathbb{N}$ different producers who sell J different products. In addition, there are $I \in \mathbb{N}$ (equally populated) groups of consumers. Each group of consumers has different preferences about different products. More specifically, let $\mathcal{J} := \{1, \dots, J\}$ be the set of producers and let $\mathcal{I} := \{1, \dots, I\}$ be the set of all possible groups. For each $i \in \mathcal{I}$ and each $j \in \mathcal{J}$, the distribution of consumers' values in group i for product j is $D_0^{ij} \in \mathcal{D}$. For one of the results below, it is further assumed that for each product $j \in \mathcal{J}$, $\{D_0^{ij}\}_{i \in \mathcal{I}}$ can be ranked by pointwise ordering. The interpretation is that for each product, different groups value a product differently and some group prefers a product more than others.

For each producer $j \in \mathcal{J}$, her marginal production cost $c_j \in C_j = [\underline{c}_j, \overline{c}_j]$ is her private information that follows a distribution G_j. Assume that the marginal costs are independent across producers. Define $C := \prod_{j \in \mathcal{J}} C_j$ and use $c = (c_1, \dots, c_J)$ to denote a typical element of C. Also, let $G := \prod_{j \in \mathcal{J}} G_j$ be the joint distribution of the producers' marginal costs. As in the baseline model, each producer can sell her product to the consumers but does not know the individual consumer's value a priori. Furthermore, the producer does not have the targeting technology and thus the consumers she faces in absence of the data broker are summarized by

$$D_0^j := \frac{1}{I} \sum_{i \in \mathcal{I}} D_0^{ij}.$$

That is, without targeting, the consumers who see producer j's product are uniformly drawn from each group.

The data broker can create market segmentations and sell them to the producers. In addition, he can help the producers target their products to different group of consumers. Formally, for any $i \in \mathcal{I}$ and any $j \in \mathcal{J}$, let $\mathcal{S}_{ij}$ denote the collection of $s \in \Delta(\mathcal{D})$ satisfying (1) with D_0 being replaced by D_0^{ij}. A mechanism is defined as a tuple $(\sigma, \tau, q) = (\sigma_{ij}, \tau_j, q_{ij})_{i \in \mathcal{I}, j \in \mathcal{J}}$, where for each $i \in \mathcal{I}$, $j \in \mathcal{J}$, $\sigma_{ij} : C \to \mathcal{S}_{ij}$ is the *segmentation scheme*; $q_{ij} : C \to [0, 1]$ such that $\sum_{i \in \mathcal{I}} q_{ij} \leq 1$ is the *targeting scheme* so that $q_{ij}(c)$ stands for the fraction of consumers of group i that can see product j;[21] and $\tau_j : C \to \mathbb{R}$ is the transfer scheme for producer j. A mechanism (σ, τ, q) is said to be incentive compatible if for any $j \in \mathcal{J}$ and for any $c_j, c_j' \in C_j$,

$$\begin{aligned} &\mathbb{E}_{c_{-j}} \left[\sum_{i \in \mathcal{I}} \int_{\mathcal{D}} \pi_D(c_j) \sigma_{ij}(\mathrm{d}D | c_j, c_{-j}) q_{ij}(c_j, c_{-j}) - \tau_j(c_j, c_{-j}) \right] \\ \geq &\mathbb{E}_{c_{-j}} \left[\sum_{i \in \mathcal{I}} \int_{\mathcal{D}} \pi_D(c_j) \sigma_{ij}(\mathrm{d}D | c_j', c_{-j}) q_{ij}(c_j', c_{-j}) - \tau_j(c_j', c_{-j}) \right], \end{aligned}$$

[21] Targeting can only re-direct the consumers who are able to see each product but cannot create new demand. As such, the total volume of consumers who can see product j must be less than $\sum_{i \in \mathcal{I}} 1/I = 1$.

and is individually rational if for any $j \in \mathcal{J}$ and any $c_j \in C_j$,

$$\mathbb{E}_{c_{-j}}\left[\sum_{i\in\mathcal{I}}\int_{\mathcal{D}}\pi_D(c_j)\sigma_{ij}(\mathrm{d}D|c_j,c_{-j})q_{ij}(c_j,c_{-j})-\tau_j(c_j,c_{-j})\right]\geq \pi_{D_0^j}(c_j).$$

Theorem 3 can be generalized to the environment in which targeted marketing is possible, as summarized in Theorem 10.

Theorem 10. *For any demands $\{D_0^{ij}\}_{i\in\mathcal{I},j\in\mathcal{J}}$ and any marginal cost distributions $\{G_j\}_{j\in\mathcal{J}}$, there exists an incentive feasible mechanism that maximizes the data broker's revenue. Moreover, under any revenue-maximizing mechanism, consumers retain zero surplus.*

Theorem 10 implies that even with the additional targeting technology, the consumers still retain no surplus. The reason is that, even though the ability to target consumers increases total surplus, the data broker can always design segmentations and targeting schemes that extract all of the additional surplus created by targeting. The groups of consumers whose values are low will not be exposed to a product, whereas the surplus of the groups of consumers whose values are high enough are entirely extracted away due to price discrimination, even if they are targeted.

In addition to the implications for consumer surplus, since every group of consumer can buy from all of the J producers as long as they see the product, the data broker's problem is in fact similar to that in the baseline model. To maximize revenue, he will simply select the most profitable group of consumers for producer j and target producer j's product to that group. This observation leads to the following generalization of Theorem 7. That is, even in environments where targeted marketing is possible, under certain appropriate assumptions about the distributions of marginal costs and the market demands, data brokership and price-controlling data brokership are still outcome-equivalent.

Theorem 11. *For any demands $\{D_0^{ij}\}_{i\in\mathcal{I},j\in\mathcal{J}}\subset\mathcal{D}$ such that $\{D_0^{ij}\}_{i\in\mathcal{I}}$ is ordered according to pointwise ordering for each $j\in\mathcal{J}$, and for any regular distributions of marginal costs $\{G_j\}_{j\in\mathcal{J}}$, suppose that for any $i\in\mathcal{I}$ and any $j\in\mathcal{J}$,$\phi_{G_j}(c)\leq\min\{\overline{\boldsymbol{p}}_{D_0^{ij}}(c),\overline{\boldsymbol{p}}_{D_0^{j}}(c)\}$ for all $c\in C$. Then data brokership is outcome-equivalent to price-controlling data brokership.*

7 Discussions

7.1 Creating Market Segmentations by Consumer Data

Throughout the paper, a market segmentation is defined as a probability measure $s\in\mathcal{S}$ that "splits" the market demand D_0 into several segments $D\in\mathcal{D}$. Although this formalization of market segmentations is well-aligned with the literature on price discrimination, a more practical way to describe a market segmentation—especially in environments where segmentations are generated by consumer data—is to define a market segment as a subset of consumer characteristics that determine the consumers' values of a product and are distributed according to some commonly known distribution. With this description, the sale of consumer data can then be interpreted as creating a partition of the characteristic set by only providing partial information about the characteristics.

Clearly, if the data broker can create other variables (e.g., purchasing propensity scores or price recommendations) that are not parts of the existing consumer characteristics, then these two modeling approaches are equivalent (see Section 4.3 for an example). However, there might be situations (due to, say, regulation or technological constraints) where the data broker cannot create additional variables but can only choose what existing characteristics to disclose and how detailed the disclosures are. In the Supplemental Material, I establish a formal result stating that these two modeling approaches are still equivalent, provided that the characteristic space is "rich enough". As a result, any market segmentation $s \in \mathcal{S}$ can be created by partitioning the underlying characteristic space and vice versa. That is, as long as the dataset contains enough of consumer characteristics, the data broker can create any $s \in \mathcal{S}$ by simply providing partial consumer data and does not need to generate additional variables.

7.2 Source of Asymmetric Information

The results in previous sections are derived under an information structure where the producer has private information about her marginal cost. Although this informational assumption captures some of the features in retail markets, it apparently does not capture all of them. Specifically, one salient informational asymmetry between a data broker and a producer in the real world is that producers often know more about how consumers' characteristics are related to their values for a particular product—perhaps due to their industry-specific knowledge that is too costly for the data broker to acquire. While optimal selling mechanisms for the data broker under this general informational environment remain an open question, the methodology developed in this paper can still provide some insights. In particular, under a parameterized information structure where the producer has private information about the market condition (as opposed to her marginal cost), all the results derived in this paper continue to hold.

More specifically, consider the following alternative information structure. There is a unit mass of consumers with unit demand for a single product. Each consumer has value $v - \xi$, where $0 < \underline{v} \leq v \leq \bar{v}$ is heterogeneous across consumers and are distributed according to $m^0 \in \Delta(V)$, while $\xi \in [0, \underline{v}]$ is the same across consumers. Both the consumers and the producer (with a commonly known marginal cost that is normalized to zero) know ξ, while the data broker only knows that ξ is drawn from a distribution G. The interpretation is that the producer knows more about the market condition (i.e., a "demand shifter" described by ξ) than the data broker does. Meanwhile, market segmentations are defined as before: A market segmentation is a probability measure $s \in \mathcal{S} \subseteq \Delta(\mathcal{D})$. It then follows that the demand in a market segment $D \in \mathcal{D}$ with market condition θ is given by $D(p+\xi)$ (i.e., $D(p+\xi)$ is the share of consumers in segment D who are willing to buy the product at price p). Under this setting, given a demand shifter θ, under any market segment $D \in \mathcal{D}$, the producer's pricing problem is given by

$$\max_{p \geq 0} pD(p+\xi),$$

which, by letting $p' = p + \xi$, is equivalent to

$$\max_{p' \geq 0} (p' - \xi)D(p') = \pi_D(\xi).$$

As a result, the informational setting above where the producer privately knows a demand shifter is equivalent to the original model where the producer has a private marginal cost ξ, and hence all the results derived above continue to hold in this alternative setting.

7.3 Policy Implications

The results above have several broader policy implications. First, in terms of welfare, although Theorem 3 implies that data brokership is undesirable for the consumers, Theorem 4 shows that the total surplus is always higher with the presence of a data broker compared with an environment where no information about the consumers' values can be disclosed. As a result, the answer to the question of whether a data broker is beneficial must depend on the objective of the policymaker and the kinds of redistributional policy tools available. If the policymaker's objective is simply maximizing total surplus, or if redistributional tools such as lump-sum transfers are available, then it is indeed beneficial to allow a data broker to sell consumer data. On the other hand, however, if the policymaker also concerns themselves with consumer surplus, and if no effective redistributional policies are available, then the presence of a data broker can be extremely unfavorable. Therefore, regarding the policy debates about whether a data broker should be allowed to collect, use and trade consumer data, it is imperative to first identify the available redistributional tools and the relative importance among consumer surplus, producer profit and total surplus.

In the case where the policymaker does wish to improve consumer surplus and no effective redistributional policies are available, Theorem 8 and Theorem 10 imply that there are limited possible policies that can be used to improve consumer surplus. Clearly, policies that help the consumers preserve some private information can improve consumer surplus. Nonetheless, according to Theorem 8, from the consumers' perspective, data brokership is still no better than all the characteristics being revealed. On the other hand, targeted marketing does not benefit the consumers either. As shown in Theorem 10, even though targeting technology can be used to further increase the total surplus, all the benefits will be extracted away from the consumers via price discrimination.

Furthermore, even if a policymaker attempts to improve consumer surplus by monitoring price discrimination, Theorem 1 implies that it is not enough to monitor only whether there is personalized pricing (i.e., first degree price discrimination). In fact, the optimal mechanisms given by Theorem 1 do not exhibit perfect price discrimination. Instead, it is a certain kind of third degree price discrimination (i.e., quasi-perfect price discrimination, induced by the quasi-perfect segmentations) that will arise in this environment. For a policymaker, prohibiting personalized pricing would not be effective in improving consumer surplus. Rather, identifying whether third degree price discrimination is operated in the form of quasi-perfect segmentation is indispensable for improving consumer surplus. Nevertheless, quasi-perfect segmentations may sometimes be difficult to identify. As illustrated in the motivating example in Section 1, this kind of segmentation can be implemented by disclosing simple consumer characteristics such as residence types. In general, it might be difficult to distinguish quasi-perfect segmentations from other basic forms of third degree price discrimination unless the policymaker has complete knowledge of the correlation between the disclosed consumer characteristics and the consumers' values. Whether it is possible to identify an quasi-perfect segmentation with less knowledge remains as a topic for future studies.

In contrast to the seemingly pessimistic implications discussed above, Theorem 5 prescribes a rather positive solution, both in terms of consumer surplus and in terms of total welfare. According to Theorem 5, if the data broker has to purchase the data from the consumers, and if the purchase takes place before the consumers learn their values, then data brokership would be Pareto-improving compared with uniform pricing. As a result, if the policymaker can establish the consumers' property right for their own data,[22] as well as a channel for the data broker to compensate the consumers, then not only the consumers can secure their surplus as if their data is not used for price discrimination (via compensation), but also the entire economy can benefit from data brokership, because less deadweight loss will be generated.

Finally, regardless of the policymaker's objective, as long as it depends only on the market outcomes, the discussions in Section 5.3 facilitates the evaluation of whether a certain market regime is desirable than another. According to Theorem 6, if the policymaker is able to eliminate the asymmetric information regarding the production cost, integrating the data

[22]For instance, just as what is stipulated by the recent regulation of the European Union, General Data Protection Regulation (GDPR, Art. 7), consumers' property right for their own data can be better protected by prohibiting all the processing of personal data unless the data subject has consented the use.

broker with the producer can be beneficial. This result is due to the fact that even when the data broker only sells consumer data to the producer, the consumer surplus is still zero. Consequently, revealing the producer's private marginal cost and encouraging vertical integration are beneficial as it does not affect the consumer surplus but eliminates all the informational frictions. In addition, the equivalence result given by Theorem 7 implies that as long as it is the producer who bears the production cost, however active the data broker is in the product market does not affect market outcomes at all. This means that, on the one hand, the data broker has no incentive to become even more active in the product market rather than only selling consumer data. In fact, together with other potential costs that are abstracted away from the model (e.g., inventory costs, shipping costs and other transaction costs), participating directly in product market can be less profitable than merely selling consumer data to the producer. On the other hand, even if the data broker does become more active in the product market, it still raises no further concerns to the policymaker. Thus, any policy intervention that prohibits the data broker entering the product market by either gaining control over prices (e.g., by establishing an online platform and allows the producer to trade on this platform while controlling the prices) or obtaining the exclusive right to (re)-sell the product would be unnecessary.[23]

8 Conclusion

In sum, this paper studies a scenario where a data broker sells consumer data and creates market segmentations. In this paper, I characterize the optimal mechanisms of the data broker and conclude that consumer surplus is always zero, that data brokership generates more total surplus than uniform pricing, and that the ability to control prices in the product market is irrelevant. Several extensions are also considered, including the case in which consumers possess some private information that cannot be disclosed and the environment where targeted marketing is available.

Several aspects can become future study topics. First, although one of the extensions of this paper considers a scenario where targeted marketing is possible, it abstracts away from the possibility that a data broker can generate "match values" between the producers and consumers. By assuming that every group of consumers can buy every product as long as they see it, the matching aspect between consumers and producers is omitted. After all, there

[23] Indeed, price-controlling data brokerpship and exclusive retail are arguably extreme benchmarks. In practice, it is likely that the producer would still be able to negotiate prices with the price-controlling data broker or to operate in the marker even after selling the product to the exclusive retailer. Nevertheless, since these more practical scenarios all generate less revenue to the data broker comparing to the extreme benchmarks, Theorem 7 still implies that the data broker would prefer only selling consumer data and not entering the product market.

is effectively no competition among the producers when there is no “scarcity” of consumers. Furthermore, the consumers’ characteristics that govern the match values can also be their private information. Second, although one of the extensions consider the case where the consumers can preserve some private information, it is restricted to certain environments. A natural direction of future research is to explore the data broker’s optimal mechanisms and their implications in a setting where the feasible market segmentation is restricted by a Blackwell upper bound. Lastly, the producer is assumed to be a product monopoly in this paper. It would be economically relevant to explore the consequences of consumer-data brokership under different industrial structures.

References

Acemoglu, D., A. Makhdoumi, A. Malekian, and A. Ozdaglar (2019): “Too Much Data: Prices and Inefficiencies in Data Market,” Working Paper.

Admati, A. R. and P. Pfleiderer (1985): “A Monopolistic Market for Information,” *Journal of Economic Theory*, 39, 400–438.

——— (1990): “Direct and Indirect Sale of Information,” *Econometrica*, 58, 901–928.

Aguirre, I., S. Cowan, and J. Vickers (2010): “Monopoly Price Discrimination and Demand Curvature,” *American Economic Review*, 100, 1601–1615.

Bergemann, D. and A. Bonatti (2015): “Selling Cookies,” *American Economic Journal: Microeconomics*, 7, 259–294.

Bergemann, D., A. Bonatti, and T. Gan (2020): “The Economics of Social Data,” Working Paper.

Bergemann, D., A. Bonatti, and A. Smolin (2018): “The Design and Price of Information,” *American Economic Review*, 108, 1–45.

Bergemann, D., B. Brooks, and S. Morris (2015): “The Limits of Price Discrimination,” *American Economic Review*, 105, 921–957.

Bergemann, D. and S. Morris (2016): “Bayes Correlated Equilibrium and the Comparison of Information Structures in Games,” *Theoretical Economics*, 11, 487–522.

Bergemann, D. and M. Pesendorfer (2007): “Information Structures in Optimal Auctions,” *Journal of Economic Theory*, 137, 580–609.

Carbajal, J. C. and J. Ely (2013): "Mechanism Design without Revenue Equivalence," *Journal of Economic Theory*, 148, 104–133.

Cowan, S. (2016): "Welfare-Increasing Third-Degree Price Discrimination," *RAND Journal of Economics*, 47, 326–340.

Dworczak, P. (2020): "Mechanism Design with Aftermarkets: Cutoff Mechanisms," *Econometrica*, Forthcoming.

Haghpanah, N. and R. Siegel (2020): "Pareto Improving Segmentation of Multi-product Markets," Working Paper.

Hart, S. and P. Reny (2019): "The Better Half of Selling Separately," *ACM Transactions on Economics and Computation*, 1, Article 18.

Ichihashi, S. (2020): "Non-Competing Data Intermediaries," Working Paper.

Jones, C. and C. Tonetti (2020): "Nonrivalry and the Economics of Data," *American Economic Review*, 110, 2819–2858.

Jullien, B. (2000): "Participation Constraints in Adverse Selection Models," *Journal of Economic Theory*, 93, 1–47.

Kamenica, E. and M. Gentzkow (2011): "Bayesian Persuasion," *American Economic Review*, 101, 2590–2615.

Lewis, T. R. and D. E. M. Sappington (1989): "Countervailing Incentives in Agency Problems," *Journal of Economic Theory*, 49, 294–313.

Maskin, E. and J. Riley (1984): "Monopoly with Incomplete Information," *RAND Journal of Economics*, 15, 171–196.

Federal Trade Commission (2014): "Data Brokers: A Call for Transparency and Accountability," https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf (accessed June 20, 2019).

Milgrom, P. and I. Segal (2002): "Envelope Theorems for Arbitrary Choice Sets," *Econometrica*, 70, 583–601.

Mussa, M. and S. Rosen (1978): "Monopoly and Product Quality," *Journal of Economic Theory*, 18, 301–307.

Myerson, R. (1979): "Incentive Compatibility and the Bargaining Problem," *Econometrica*, 47, 61–73.

——— (1981): "Optimal Auction Design," *Mathematics of Operations Research*, 6, 58–73.

Mylovanov, T. and T. Tröger (2014): "Mechanism Design by an Informed Principal: The Quasi-Linear Private-Values Case," *Review of Economic Studies*, 81, 1668–1707.

Pavan, A., I. Segal, and J. Toikka (2014): "Dynamic Mechanism Design: A Myersonian Approach," *Econometrica*, 82, 601–653.

Porter, J. E. (2005): "Helly's Selection Principle for Functions with Bounded p-Variation," *Rocky Mountain Journal of Mathematics*, 35, 675–679.

Riley, J. and R. Zeckhauser (1983): "Optimal Selling Strategies: When to Haggle, When to Hold Firm," *Quarterly Journal of Economics*, 98, 267–289.

Rochet, J.-C. (1987): "A Necessary and Sufficient Condition for Rationalizability in a Quasi-linear Context," *Journal of Mathematical Economics*, 16, 191–200.

Rockafellar, R. T. (1970): *Convex Analysis*, Princeton, NJ: Princeton University Press.

Segura-Rodriguez, C. (2020): "Selling Data," Working Paper.

Varian, H. R. (1985): "Price Discrimination and Social Welfare," *American Economic Review*, 75, 870–875.

Wei, D. and B. Green (2020): "(Reverse) Price Discrimination with Information Design," Working Paper.

Yamashita, T. (2017): "Optimal Public Information Disclosure by Mechanism Designer," Working Paper.

Yang, K. H. (2019): "Equivalence in Business Models of Informational Intermediary," Working Paper.

Appendix

A Notational Conventions

Below I first discuss more formally about the properties of the set $\mathcal{D}$. Recall that $\mathcal{D} = \mathcal{D}([\underline{v}, \overline{v}])$ is the collection of nonincreasing and left-continuous functions D on $[\underline{v}, \overline{v}]$ such that $D(\underline{v}) = 1$ and $D(\overline{v}^+) = 0$. Since for every $D \in \mathcal{D}$, there exists a unique probability measure $m^D \in \Delta(V)$ such that $D(p) = m^D(\{v \geq p\})$ for all $p \in V$, I define the topology on $\mathcal{D}$ by the following notion of convergence: For any $\{D_n\} \subseteq \mathcal{D}$ and any $D \in \mathcal{D}$, $\{D_n\} \to D$ if and only if for any bounded continuous function $f : V \to \mathbb{R}$,

$$\lim_{n\to\infty} \int_V f(v) m^{D_n}(\mathrm{d}v) = \int_V f(v) m^D(\mathrm{d}v).$$

This would corresponds to the weak-* topology on $\Delta(V)$ and hence this topology on $\mathcal{D}$ is also called the weak-* topology. As a result, $\mathcal{D}$ is a Polish space. Furthermore, notice that under this topology, $\{D_n\} \to D$ if and only if $\{D_n(p)\} \to D(p)$ for all $p \in V$ at which D is continuous.

Now I introduce some more notational conventions that are implicitly used in the main text and will be used throughout the proof. For any measurable sets X and Y, the collection of measurable functions $f : X \to Y$ is denoted by X^Y. Moreover, for any $f \in \mathbb{R}^X$, define f^+ by $f^+(x) := \max\{f(x), 0\}$ for all $x \in X$. For any $f, g \in \mathbb{R}^X$ and for any $\alpha, \beta \in \mathbb{R}$, define $\alpha f + \beta g \in \mathbb{R}^X$ by $[\alpha f + \beta g](x) := \alpha f(x) + \beta g(x)$ for all $x \in X$. If $X \subset \mathbb{R}^n$ and the partial derivative of f with respect to x_i exists for some $i \in \{1, \dots, n\}$, use

$$f_i(x_1, \dots, x_n) := \frac{\partial}{\partial x_i} f(x_1, \dots, x_n)$$

to denote this partial derivative. When $X \subseteq \mathbb{R}$, for any $x \in \mathrm{int}(X)$, let

$$f(x^+) := \lim_{x' \downarrow x} f(x') \text{ and } f(x^-) := \lim_{x' \uparrow x} f(x')$$

be the right and the left limits of f at x provided they exist, respectively.

For any measurable space X, let $\Delta^f(X) \subseteq \Delta(X)$ be the collection of probability measures on X that have finite support. The collection of probability measures $\Delta(X)$ is endowed with the algebraic structure so that for any $\mu_1, \mu_2 \in \Delta(X)$,

$$[\lambda\mu_1 + (1-\lambda)\mu_2](A) := \lambda\mu_1(A) + (1-\lambda)\mu_2(A),$$

for any measurable $A \subseteq X$. Furthermore, for any $x \in X$, $\delta_{\{x\}}$ denotes the Dirac measure that assigns probability 1 to the element x (whenever $\{x\}$ is measurable). For any probability measure $\mu \in \Delta^f(X)$ with finite support and for any measurable set $A \subseteq X$, the simplifying notation

$$\sum_{x\in A} \mu(x) := \sum_{x \in A \cap \mathrm{supp}(\mu)} \mu(\{x\})$$

will be used.

Finally, for any $D \in \mathcal{D}$, let $\mathcal{S}_D$ denote the collection of $s \in \Delta(\mathcal{D})$ such that (1) holds with D_0 being replaced by D (so that $\mathcal{S}_{D_0} = \mathcal{S}$). Also, let D^{-1} denote the inverse demand of D, where D^{-1} is defined as

$$D^{-1}(q) := \sup\{p \in V : D(p) \geq q\}, \; \forall q \in [0,1]. \tag{14}$$

B Technical Lemmas and Proofs

This section contains technical lemmas that establish continuities of several critical objects. These continuity results are crucial for proving the main results.

Lemma 5. *For any $D \in \mathcal{D}$, $\pi_D \in \mathbb{R}_+^C$ is continuous and convex. Furthermore, for any $\boldsymbol{p} \in \boldsymbol{P}$, and for any $c \in C$, $-D(\boldsymbol{p}_D(c))$ is a subgradient of π_D at c. In particular, for any $\underline{c} \leq c < c' \leq \overline{c}$,*

$$\pi_D(c) - \pi_D(c') = \int_c^{c'} D(\boldsymbol{p}_D(z))\,\mathrm{d}z \tag{15}$$

for any $\boldsymbol{p} \in \boldsymbol{P}$

Proof. By definition of π_D, for any $c \in C$,

$$\pi_D(c) = \max_{p \in \mathbb{R}_+}(p - c)D(p).$$

As such, π_D is convex for all $D \in \mathcal{D}$ since it is the pointwise supremum of a family of affine functions. Moreover, for any $\boldsymbol{p} \in \boldsymbol{P}$ and for any $c, c' \in C$,

$$\begin{aligned}
0 \leq & \pi_D(c') - (\boldsymbol{p}_D(c) - c')D(\boldsymbol{p}_D(c)) \\
= & \pi_D(c') + c'D(\boldsymbol{p}_D(c)) - \boldsymbol{p}_D(c)D(\boldsymbol{p}_D(c)) \\
= & \pi_D(c') - cD(\boldsymbol{p}_D(c)) + c'D(\boldsymbol{p}_D(c)) - (\boldsymbol{p}_D(c) - c)D(\boldsymbol{p}_D(c)) \\
= & \pi_D(c') - [-D(\boldsymbol{p}_D(c))(c' - c) + \pi_D(c)].
\end{aligned}$$

Thus, $-D(\boldsymbol{p}_D(c))$ is a subgradient of π_D at c. Together with convexity of π_D, π_D is differentiable almost everywhere and

$$\pi_D'(c) = -D(\boldsymbol{p}_D(c)),$$

for almost all $c \in C$. Thus, since π_D is convex, for any $\underline{c} \leq c < c' \leq \overline{c}$,

$$\pi_D(c) - \pi_D(c') = \int_c^{c'} D(\boldsymbol{p}_D(z))\,\mathrm{d}z,$$

for any $\boldsymbol{p} \in \boldsymbol{P}$.

For continuity, notice that for any $c \in C$,

$$\pi_D(c) = \max_{(p,q) \in \Xi}(p - c)q,$$

where $\Xi := \mathrm{cl}(\{(p, D(p)) : p \in V\})$ is a compact set in $\mathbb{R}^2$. Therefore, by Berge's theorem of maximum, π_D is continuous on C. ∎

Lemma 6. *The correspondence $\boldsymbol{P}$ is compact-valued and thus*

$$\overline{\boldsymbol{p}}_D(c) := \max \boldsymbol{P}_D(c)$$

and

$$\underline{\boldsymbol{p}}_D(c) := \min \boldsymbol{P}_D(c)$$

are well-defined for all $c \in C$, $D \in \mathcal{D}$. Furthermore, for any $D \in \mathcal{D}$, the correspondence $\boldsymbol{P}_D : C \rightrightarrows \mathbb{R}_+$ is upper-hemicontinuous. In particular, $\overline{\boldsymbol{p}}_D \in \mathbb{R}_+^C$ is right-continuous and $\underline{\boldsymbol{p}}_D \in \mathbb{R}_+^C$ is left-continuous.

Proof. Consider any $c \in C$ and $D \in \mathcal{D}$. Suppose that $\{p_n\} \subseteq \boldsymbol{P}_D(c)$ and $\{p_n\} \to p$ for some $p \in \mathbb{R}_+$. Since the function $p \mapsto (p-c)D(p)$ is upper-semicontinuous,

$$\pi_D(c) = \limsup_{n\to\infty}(p_n - c)D(p_n) \leq (p-c)D(p) \leq \pi_D(c).$$

Thus, $p \in \boldsymbol{P}_D(c)$. As a result, since $\boldsymbol{P}_D(c) \subseteq \overline{V}$ (see footnote 7), for all $c \in C$ and $D \in \mathcal{D}$, $\boldsymbol{P}_D(c)$ is a closed subset of a compact set, which implies that $\overline{\boldsymbol{p}}_D(c)$ and $\underline{\boldsymbol{p}}_D(c)$ are well-defined.

Now consider any $D \in \mathcal{D}$. To show upper-hemicontinuity of $\boldsymbol{P}_D$, it suffices to show that for any sequences $\{c_n\} \subseteq C$ and $\{p_n\} \subseteq \mathbb{R}_+$ such that $\{p_n\} \to p \in \mathbb{R}_+$ and $\{c_n\} \to c \in C$ and that $p_n \in \boldsymbol{P}_D(c_n)$ for all $n \in \mathbb{N}$, $p \in \boldsymbol{P}_D(c)$. Indeed, for any $n \in \mathbb{N}$, since $p_n \in \boldsymbol{P}_D(c_n)$, $\pi_D(c_n) = (p_n - c_n)D(p_n)$ for all $n \in \mathbb{N}$. Moreover, since $\pi_D \in \mathbb{R}_+^C$ according to Lemma 5,

$$\lim_{n\to\infty} \pi_D(c_n) = \pi_D(c).$$

Therefore, since D is upper-semicontinuous,

$$\pi_D(c) = \lim_{n\to\infty} \pi_D(c_n) = \limsup_{n\to\infty}(p_n - c_n)D(p_n) \leq (p-c)D(p) \leq \pi_D(c).$$

Thus, $p \in \boldsymbol{P}_D(c)$ as desired. Finally, since for any $\boldsymbol{p} \in \boldsymbol{P}$, $\boldsymbol{p}_D \in \mathbb{R}_+^C$ is nondecreasing, upper-hemicontinuity of $\boldsymbol{p}_D$ then implies right-continuity of $\overline{\boldsymbol{p}}_D$ and left-continuity of $\underline{\boldsymbol{p}}_D$. This completes the proof. ■

Lemma 7. *For any $D \in \mathcal{D}$, the function $c \mapsto D(\overline{\boldsymbol{p}}_D(c))$ is right-continuous.*

Proof. Consider any $D \in \mathcal{D}$ and any $c \in C$. By Lemma 6,

$$\lim_{c' \downarrow c} \overline{\boldsymbol{p}}_D(c') = \overline{\boldsymbol{p}}_D(c).$$

Together with continuity of π_D, which is due to Lemma 5,

$$\begin{aligned}
(\overline{\boldsymbol{p}}_D(c) - c)D(\overline{\boldsymbol{p}}_D(c)) =& \pi_D(c) \\
=& \lim_{c' \downarrow c} \pi_D(c') \\
=& \lim_{c' \downarrow c} (\overline{\boldsymbol{p}}_D(c') - c')D(\overline{\boldsymbol{p}}_D(c')) \\
=& (\overline{\boldsymbol{p}}_D(c) - c) \cdot \lim_{c' \downarrow c} D(\overline{\boldsymbol{p}}_D(c')),
\end{aligned}$$

and hence

$$\lim_{c' \downarrow c} D(\overline{\boldsymbol{p}}_D(c')) = D(\overline{\boldsymbol{p}}_D(c)),$$

as desired. ■

Lemma 8. *For any $c \in C$, the function $D \mapsto \pi_D(c)$ is continuous on $\mathcal{D}$.*

Proof. Since $V \subseteq \mathbb{R}_+$ is bounded, this lemma is a special case of Theorem 12 of Hart and Reny (2019) when $k = 1$. ■

Lemma 9. *For any $c \in (\underline{c}, \overline{c})$, the function $D \mapsto D(\overline{\boldsymbol{p}}_D(c))$ is lower-semicontinuous on $\mathcal{D}$.*

Proof. For any $c \in (\underline{c}, \overline{c})$ and for any $D \in \mathcal{D}$, define $\pi'_D(c^+)$ as

$$\pi'_D(c^+) := \lim_{c' \downarrow c} \frac{\pi_D(c') - \pi_D(c)}{c' - c}.$$

Since π_D is convex, $\pi'_D(c^+)$ is well-defined. Furthermore, by Lemma 5, $-D(\overline{\boldsymbol{p}}_D(c))$ is a subgradient of π_D at c and therefore, for any $c' > c$,

$$\frac{\pi_D(c') - \pi_D(c)}{c' - c} \geq -D(\overline{\boldsymbol{p}}_D(c)),$$

which implies that

$$\pi'_D(c^+) \geq D(\overline{\boldsymbol{p}}_D(c)). \tag{16}$$

Meanwhile, by (15), for any $c' > c$,

$$\frac{\pi_D(c') - \pi_D(c)}{c' - c} = \frac{1}{c' - c} \int_c^{c'} -D(\overline{\boldsymbol{p}}_D(z))\, \mathrm{d}z \leq -D(\overline{\boldsymbol{p}}_D(c')).$$

Thus, by Lemma 7,

$$\pi'_D(c^+) \leq \lim_{c' \downarrow c} -D(\overline{\boldsymbol{p}}_D(c')) = -D(\overline{\boldsymbol{p}}_D(c)). \tag{17}$$

Combining (16) and (17),

$$\pi'_D(c^+) = D(\overline{\boldsymbol{p}}_D(c)).$$

Now consider any $D \in \mathcal{D}$ and any $\{D_n\} \subseteq \mathcal{D}$ such that $\{D_n\} \to D$, Lemma 8 implies that $\{\pi_{D_n}\} \to \pi_D$ pointwise. Thus, for any $c \in (\underline{c}, \overline{c})$, by Theorem 24.5 of Rockafellar (1970),

$$-\liminf_{n \to \infty} D_n(\overline{\boldsymbol{p}}_{D_n}(c)) = \limsup_{n \to \infty} \pi'_{D_n}(c^+) \leq \pi'_D(c^+) = -D(\overline{\boldsymbol{p}}_D(c)).$$

Therefore, for any $c \in (\underline{c}, \overline{c})$,

$$\liminf_{n \to \infty} D_n(\overline{\boldsymbol{p}}_{D_n}(c)) \geq D(\overline{\boldsymbol{p}}_D(c)),$$

as desired. ∎

Lemma 10. *For any $c \in C$, the function $D \mapsto \overline{\boldsymbol{p}}_D(c)$ is upper-semicontinuous on $\mathcal{D}$.*

Proof. Consider any $c \in C$ and any sequence $\{D_n\} \subseteq \mathcal{D}$ such that $\{D_n\} \to D$ for some $D \in \mathcal{D}$. Let

$$\bar{p} := \limsup_{n \to \infty} \overline{\boldsymbol{p}}_{D_n}(c).$$

Take a subsequence $\{D_{n_k}\} \subseteq \{D_n\}$ such that

$$\lim_{k \to \infty} \overline{\boldsymbol{p}}_{D_{n_k}}(c) = \bar{p}.$$

First notice that since $D \in \mathcal{D}$ is upper-semicontinuous, for any sequence $\{\delta_k\} \subset \mathbb{R}_+$ such that $\{\delta_k\} \to 0$,

$$\limsup_{k \to \infty} (\overline{\boldsymbol{p}}_{D_{n_k}}(c) - c) D_{n_k}(\overline{\boldsymbol{p}}_{D_{n_k}}(c) - \delta_k) \leq (\bar{p} - c) D(\bar{p}). \tag{18}$$

Moreover, by the definition of the Lévy Prokhorov metric, for any $k \in \mathbb{N}$,

$$D_{n_k}(\overline{\boldsymbol{p}}_{D_{n_k}}(c)) \leq D\left(\overline{\boldsymbol{p}}_{D_{n_k}}(c) - \left(\rho(D_{n_k}, D) + \frac{1}{k}\right)\right) - \left(\rho(D_{n_k}, D) + \frac{1}{k}\right), \tag{19}$$

where $\rho : \mathcal{D} \times \mathcal{D} \to \mathbb{R}_+$ is the Lévy Prokhorov metric. Together, since $\{\rho(D_{n_k}, D)\} \to 0$ as $k \to \infty$, which is because $\{D_{n_k}\} \to D$ as $k \to \infty$, we have

$$\begin{aligned}
\pi_D(c) &= \lim_{k\to\infty} \pi_{D_{n_k}}(c) \\
&= \lim_{k\to\infty} (\overline{\boldsymbol{p}}_{D_{n_k}}(c) - c) D_{n_k}(\overline{\boldsymbol{p}}_{D_{n_k}}(c)) \\
&\leq \limsup_{k\to\infty} (\overline{\boldsymbol{p}}_{D_{n_k}}(c) - c) \left[D\left(\overline{\boldsymbol{p}}_{D_{n_k}}(c) - \left(\rho(D_{n_k}, D) + \frac{1}{k} \right) \right) - \left(\rho(D_{n_k}, D) + \frac{1}{k} \right) \right] \\
&\leq (\bar{p} - c) D(\bar{p}) \\
&\leq \pi_D(c),
\end{aligned}$$

where the first equality follows from Lemma 8, the first inequality follows from (19), and the second inequality follows from $\{\rho(D_{n_k}, D)\} \to 0$ as $k \to \infty$, (18), as well as the fact that $\overline{\boldsymbol{p}}_{D_{n_k}}(c) \leq \max \bar{V}$ (see footnote 7). As a result, it then follows that $\bar{p} \in \boldsymbol{P}_D(c)$ and therefore $\bar{p} \leq \overline{\boldsymbol{p}}_D(c)$. Thus,

$$\limsup_{n\to\infty} \overline{\boldsymbol{p}}_{D_n}(c) = \bar{p} \leq \boldsymbol{p}_D(c),$$

as desired. ∎

C Crucial Properties of Quasi-Perfect Schemes

This section summarizes some crucial properties of a ψ-quasi-perfect scheme.

Lemma 11. *Consider any nondecreasing function $\psi \in \mathbb{R}_+^C$ with $c \leq \psi(c)$ for all $c \in C$. Suppose that for any $c \in C$, $\sigma(c) \in \mathcal{S}$ is a $\psi(c)$-quasi-perfect segmentation for c. Then,*

1. *$\int_{\mathcal{D}} D(p)\sigma(\mathrm{d}D|c) = D_0(p)$ for all $p \in V$ and for all $c \in C$.*
2. *$\sigma : C \to \Delta(\mathcal{D})$ is measurable.*
3. *$\int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(c))\sigma(\mathrm{d}D|c) = D_0(\psi(c))$ for all $c \in C$.*
4. *$\int_{\mathcal{D}} \overline{\boldsymbol{p}}_D(c) D(\overline{\boldsymbol{p}}_D(c))\sigma(\mathrm{d}D|c) = \int_{\{v \geq \psi(c)\}} v D_0(\mathrm{d}v)$ for all $c \in C$.*

Proof. For any nondecreasing function $\psi \in \mathbb{R}_+^C$ with $c \leq \psi(c)$ for all $c \in C$, since for any $c \in C$, $\sigma(c) \in \mathcal{S}$ is a $\psi(c)$-quasi-perfect segmentation for c, by definition,

$$\int_{\mathcal{D}} D(p)\sigma(\mathrm{d}D|c) = D_0(p), \tag{20}$$

for all $p \in V$, which proves assertion 1. Furthermore, since ψ is nondecreasing and is thus continuous except at countably many points, $\sigma : C \to \Delta(\mathcal{D})$ is measurable, which establishes assertion 2. For assertion 3, notice that for any $c \in C$, since $\sigma(c) \in \mathcal{S}$ is a $\psi(c)$-quasi-perfect segmentation for c, for any $D \in \mathrm{supp}(\sigma(c))$ such that $D(\overline{\boldsymbol{p}}_D(c)) > 0$,

$$D(\overline{\boldsymbol{p}}_D(c)) = D(\max(\mathrm{supp}(D))) = D(\psi(c))$$

and thus

$$\int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(c))\sigma(\mathrm{d}D|c) = \int_{\mathcal{D}} D(\psi(c))\sigma(\mathrm{d}D|c) = D_0(\psi(c)),$$

where the last equality follows from (20). This proves assertion 3. Finally, to prove assertion 4, consider any $c \in C$. First notice that if $D_0(c) = 0$, then assertion 4 clearly holds as both sides would be zero. Now suppose that $D_0(c) > 0$. The fact that $\sigma(c) \in \mathcal{S}$ is a $\psi(c)$-quasi-perfect segmentation for c ensures that $D_0(\psi(c)) > 0$. Then, for any $v \in [\psi(c), \overline{v}]$, let

$$H(v) := \sigma(\{D \in \mathcal{D} : \max(\text{supp}(D)) \leq v\}|c).$$

Since $\sigma(c)$ is a probability measure, H is nondecreasing and right-continuous and hence induces a Borel measure μ_H on $[\psi(c), \overline{v}]$. On the other hand, for any measurable sets $A, B \subseteq [\psi(c), \overline{v}]$, define

$$K(A|B) := \int_{\{D \in \mathcal{D}:\max(\text{supp}(D)) \in A\}} m^D(B)\sigma(\mathrm{d}D|c).$$

Notice that for any measurable set $B \subseteq [\psi(c), \overline{v}]$, $K(\cdot|B)$ is a measure and is absolutely continuous with respect to μ_H and hence there exists a (essentially) unique Radon-Nikodym derivative $v \mapsto m^v(B)$ such that for any measurable $A \subseteq [\psi(c), \overline{v}]$,

$$K(A|B) = \int_{v \in A} m^v(B)H(\mathrm{d}v). \tag{21}$$

In particular, by definition of K and by (20), for any measurable set $B \subseteq [\psi(c), \overline{v}]$,

$$\int_{[\psi(c),\overline{v}]} m^v(B)H(\mathrm{d}v) = K([\psi(c), \overline{v}]|B) = \int_{\mathcal{D}} m^D(B)\sigma(\mathrm{d}D|c) = m^0(B). \tag{22}$$

Moreover, since for any measurable set $A \subseteq [\psi(c), \overline{v}]$, $K(A|\cdot)$ is a measure on $[\psi(c), \overline{v}]$ and thus m^v is also a measure on $[\psi(c, \overline{v})]$ for μ_H-almost all $v \in [\psi(c), \overline{v}]$. Furthermore, since $\sigma(c) \in \mathcal{S}$ is a $\psi(c)$-quasi-perfect segmentation for c, for any measurable sets $A, B \subseteq [\psi(c), \overline{c}]$,

$$K(A|B) = m^0(A \cap B) = K(B|A)$$

and hence, for any measurable sets $A, B \subseteq [\psi(c), \overline{v}]$,

$$\int_A m^v(B)H(\mathrm{d}v) = \int_B m^v(A)H(\mathrm{d}v). \tag{23}$$

As a result,

$$\begin{aligned}
\int_{\mathcal{D}} \overline{\boldsymbol{p}}_D(c)D(\overline{\boldsymbol{p}}_D(c))\sigma(\mathrm{d}D|c) &= \int_{\mathcal{D}} \overline{\boldsymbol{p}}_D(c)D(\psi(c))\sigma(\mathrm{d}D|c) \\
&= \int_{\mathcal{D}} \max(\text{supp}(D))m^D([\psi(c), \overline{v}])\sigma(\mathrm{d}D|c) \\
&= \int_{[\psi(c),\overline{v}]} vK(\mathrm{d}v|[\psi(c), \overline{v}]) \\
&= \int_{[\psi(c),\overline{v}]} vm^v([\psi(c), \overline{v}])H(\mathrm{d}v) \\
&= \int_{v \in [\psi(c),\overline{v}]} \int_{v' \in [\psi(c),\overline{v}]} vm^v(\mathrm{d}v')H(\mathrm{d}v) \\
&= \int_{v \in [\psi(c),\overline{v}]} v \left(\int_{v' \in [\psi(c),\overline{v}]} m^{v'}(\mathrm{d}v)H(\mathrm{d}v') \right) \\
&= \int_{[\psi(c),\overline{v}]} vD_0(\mathrm{d}v),
\end{aligned}$$

where the second equality follows from the fact that $\sigma(c)$ is a $\psi(c)$-quasi-perfect segmentation for c, the third equality follows from the definition of K, the fourth equality follows from (21), the sixth equality follows from (23), and the last equality follows from (22). This completes the proof. ■

Lemma 12. *For any nondecreasing function $\psi \in \mathbb{R}_+^C$ with $c \leq \psi(c)$ for all $c \in C$, suppose that $\sigma \in \mathcal{S}^C$ is a ψ-quasi-perfect scheme. Then for any $c, c' \in C$ with $c < c'$,*

$$\int_c^{c'} \left(\int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(z))(\sigma(\mathrm{d}D|c) - \sigma(\mathrm{d}D|z)) \right) \mathrm{d}z \geq 0.$$

Proof. Consider any $c, c' \in C$ such that $c < c'$. Notice that since $\sigma \in \mathcal{S}^C$ is a ψ-quasi-perfect scheme, by Lemma 11,

$$\int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(z))\sigma(\mathrm{d}D|z) = D_0(\psi(z)).$$

In addition, since $\sigma(c)$ is a $\psi(c)$-quasi-perfect segmentation for c, for any $z > c$ and for any $D \in \mathrm{supp}(\sigma(c))$, if $D(c) > 0$ and $\max(\mathrm{supp}(D)) \geq z$, then $\overline{\boldsymbol{p}}_D(z) = \overline{\boldsymbol{p}}_D(c)$. On the other hand, if $D(c) > 0$ and $\max(\mathrm{supp}(D)) < z$, then $D(\overline{\boldsymbol{p}}_D(z)) = 0$. Also, notice that $D(c) = 0$ implies $D(z) = 0$. Together, if $z \leq \psi(c)$,

$$\int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(z))\sigma(\mathrm{d}D|c) = \int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(c))\sigma(\mathrm{d}D|c) = D_0(\psi(c))$$

and if $z > \psi(c)$,

$$\begin{aligned}
\int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(z))\sigma(\mathrm{d}D|c) &= \int_{\{D \in \mathcal{D}: \max(\mathrm{supp}(D)) \geq z\}} D(\overline{\boldsymbol{p}}_D(c))\sigma(\mathrm{d}D|c) \\
&= \int_{\mathcal{D}} D(z)\sigma(\mathrm{d}D|c) \\
&= D_0(z).
\end{aligned}$$

As a result,

$$\begin{aligned}
&\int_c^{c'} \left(\int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(z))(\sigma(\mathrm{d}D|c) - \sigma(\mathrm{d}D|z)) \right) \mathrm{d}z \\
= &\int_c^{c'} \left(\int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(z))\sigma(\mathrm{d}D|c) - \int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(z))\sigma(\mathrm{d}D|z) \right) \mathrm{d}z \\
= &\int_c^{c'} (\min\{D_0(\psi(c)), D_0(z)\} - D_0(\psi(z)))\, \mathrm{d}z \\
\geq &0,
\end{aligned}$$

where the inequality follows from the fact that $z \leq \psi(z)$ and $\psi(c) \leq \psi(z)$ for all $z \in [c, c']$, which in turn relies on the hypothesis that $c \leq \psi(c)$ for all $c \in C$ and that ψ is nondecreasing. This completes the proof. ■

Lemma 13. *Consider any function $\psi \in \mathbb{R}_+^C$ with $c \leq \psi(c)$ for all $c \in C$. Given any $\{D_n\} \subset \mathcal{D}$ and $\{\sigma_n\} \subset \mathcal{S}_{D_n}^C$. Suppose that $\{\sigma_n\} \to \sigma$ pointwise and $\{D_n\} \to D_0$ for some $\sigma \in \Delta(\mathcal{D})^C$ and $D_0 \in \mathcal{D}$. Then $\sigma \in \mathcal{S}_{D_0}^C$. Moreover, suppose further that σ_n is a ψ-quasi-perfect scheme for all $n \in \mathbb{N}$. Then σ is a ψ-quasi-perfect scheme.*

Proof. First notice that since $\sigma_n \in \mathcal{S}^C_{D_n}$ is a ψ-quasi-perfect scheme, Lemma 11 ensures that σ_n is measurable. Then, since $\{\sigma_n\} \to \sigma$ pointwise, σ is also measurable. Moreover, since $\{D_n\} \to D_0$ and $\{\sigma_n\} \to \sigma$, for any bounded continuous function $f : V \to \mathbb{R}$ and for any $c \in C$

$$\begin{aligned}\int_V f(v)\left(\int_{\mathcal{D}} D(\mathrm{d}v)\sigma(\mathrm{d}D|c)\right) &= \int_{\mathcal{D}}\left(\int_V f(v)D(\mathrm{d}v)\right)\sigma(\mathrm{d}D|c)\\ &= \lim_{n\to\infty}\int_{\mathcal{D}}\left(\int_V f(v)D(\mathrm{d}v)\right)\sigma_n(\mathrm{d}D|c)\\ &= \lim_{n\to\infty}\int_V f(v)\left(\int_{\mathcal{D}} D(\mathrm{d}v)\sigma(\mathrm{d}D|c)\right)\\ &= \lim_{n\to\infty}\int_V f(v)D_n(\mathrm{d}v)\\ &= \int_V f(v)D_0(\mathrm{d}v),\end{aligned}$$

where the first and the third equality follow from interchanging the order of integrals, the second equality follows from the fact that the integrand in the parentheses is a bounded continuous function of D and from weak-*convergence of $\{\sigma_n(c)\}$, the fourth equality is due to the fact that $\sigma_n(c) \in \mathcal{S}_{D_n}$, and the last equality follows from the weak-* convergence of $\{D_n\}$. Thus, by the Riesz representation theorem,

$$\int_{\mathcal{D}} D(p)\sigma(\mathrm{d}D|c) = D_0(p),\ \forall p \in V,\ c \in C.$$

This proves that $\sigma \in \mathcal{S}^C$.

Now suppose that σ_n is a ψ-quasi-perfect scheme for all $n \in \mathbb{N}$ and suppose that, by way of contradiction, $\sigma \in \mathcal{S}^C$ is not a ψ-quasi-perfect scheme. Then there exists a positive G-measure of c and a positive $\sigma(c)$-measure of $D \in \mathcal{D}$ such that $D(\{v > c\}) > 0$ and either $\#\{v \in \mathrm{supp}(D) : v \geq \psi(c)\} \neq 1$ or $\max(\mathrm{supp}(D)) \notin \boldsymbol{P}_D(c)$ (i.e., $D(\overline{\boldsymbol{p}}_D(c)) > 0$). As such, there is a positive G-measure of c and a positive $\sigma(c)$-measure of D such that

$$\begin{aligned}\int_{\{v\geq\psi(c)\}}(v-\psi(c))D(\mathrm{d}v) \geq& \int_{\{v\geq\overline{\boldsymbol{p}}_D(c)\}}(v-\psi(c))D(\mathrm{d}v)\\ =&(\overline{\boldsymbol{p}}_D(c)-\psi(c))D(\overline{\boldsymbol{p}}_D(c)) + \int_{\{v\geq\overline{\boldsymbol{p}}_D(c)\}}(v-\overline{\boldsymbol{p}}_D(c))D(\mathrm{d}v)\\ \geq&(\overline{\boldsymbol{p}}_D(c)-\psi(c))D(\overline{\boldsymbol{p}}_D(c)),\end{aligned}$$

with at least one inequality being strict. Thus, there exists a positive G-measure of $c \in C$ such that

$$\int_{\mathcal{D}}(\overline{\boldsymbol{p}}_D(c)-\psi(c))D(\overline{\boldsymbol{p}}_D(c))\sigma(\mathrm{d}D|c) < \int_V (v-\psi(c))^+ D_0(\mathrm{d}v).$$

However, by Lemma 8 and Lemma 9, for Lebesgue almost all $c \in C$,

$$\begin{aligned}
&\int_{\mathcal{D}} (\overline{\boldsymbol{p}}_D(c) - \psi(c)) D(\overline{\boldsymbol{p}}_D(c)) \sigma(\mathrm{d}D|c) \\
&= \int_{\mathcal{D}} \pi_D(c) \sigma(\mathrm{d}D|c) - (\psi(c) - c) \int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(c)) \sigma(\mathrm{d}D|c) \\
&\geq \lim_{n\to\infty} \int_{\mathcal{D}} \pi_D(c) \sigma_n(\mathrm{d}D|c) - \liminf_{n\to\infty} (\psi(c) - c) \int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(c)) \sigma_n(\mathrm{d}D|c) \\
&= \limsup_{n\to\infty} \left[\int_{\mathcal{D}} \pi_D(c) \sigma_n(\mathrm{d}D|c) - (\psi(c) - c) \int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(c)) \sigma_n(\mathrm{d}D|c) \right] \qquad (24) \\
&= \limsup_{n\to\infty} \int_V (v - \psi(c))^+ D_n(\mathrm{d}v) \\
&= \lim_{n\to\infty} \int_V (v - \psi(c))^+ D_n(\mathrm{d}v) \\
&= \int_V (v - \psi(c))^+ D_0(\mathrm{d}v),
\end{aligned}$$

a contradiction. Here, the first inequality follows from the fact that $\{\sigma_n(c)\} \to \sigma(c)$, Lemma 8 and Lemma 9; the second equality follows from the properties of the lim inf and lim sup operators;[24] the third equality follows from the fact that $\sigma_n(c) \in \mathcal{S}_{D_n}$ and is a $\psi(c)$-quasi-perfect segmentation for c; and the last two equalities follow from the fact that the function $(v - \psi(c))^+$ is bounded and continuous in v and that $\{D_n\} \to D_0$. Therefore, σ must be a ψ-quasi-perfect scheme. ■

D Proofs for Optimal Mechanisms

D.1 Proof of Proposition 2

In this section, I first prove Proposition 2 and obtain an upper bound for the data broker's revenue. That is, I first solve the relaxed problem where the prices are also contractable. To this end, I first introduce the revenue-equivalence formula for the price-controlling data broker.

Lemma 14. *For the price-controlling data broker, a mechanism $(\sigma, \tau, \boldsymbol{\gamma})$ is incentive compatible if and only if*

1. *There exists some $\bar{\tau} \in \mathbb{R}$ such that for any $c \in C$,*

$$\tau(c) = \int_{\mathcal{D}} \int_{\mathbb{R}_+} (p - c) D(p) \boldsymbol{\gamma}(\mathrm{d}p|D, c) \sigma(\mathrm{d}D|c) - \int_c^{\bar{c}} \int_{\mathcal{D}} \int_{\mathbb{R}_+} D(p) \boldsymbol{\gamma}(\mathrm{d}p|D, z) \sigma(\mathrm{d}D|z)\, \mathrm{d}z - \bar{\tau}.$$

[24] More precisely, this follows from the following properties: For any real sequences $\{a_n\}$, $\{b_n\}$,

$$-\liminf_{n\to\infty} b_n = \limsup_{n\to\infty} (-b_n).$$

Moreover, if $\{a_n\}$ is convergent, then

$$\limsup_{n\to\infty} (a_n + b_n) = \lim_{n\to\infty} a_n + \limsup_{n\to\infty} b_n.$$

2. *The function*

$$c \mapsto \int_{\mathcal{D}} \int_{\mathbb{R}_+} D(p)\boldsymbol{\gamma}(\mathrm{d}p|D,c)\sigma(\mathrm{d}D|c)$$

is nonincreasing.

Proof. For necessity, consider any incentive compatible mechanism $(\sigma, \tau, \boldsymbol{\gamma})$. Let

$$u(c,c') := \int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} (p-c)D(p)\boldsymbol{\gamma}(\mathrm{d}p|D,c') \right) \sigma(\mathrm{d}D|c') - \tau(c'),\ \forall c, c' \in C$$

denote the producer's net profit when her marginal cost is c and reports c'. By incentive compatibility, for any $c \in C$

$$U(c) := u(c,c) \geq u(c,c'),$$

Since $u_1(c,c') = -\int_{\mathcal{D}} \int_{\mathbb{R}_+} D(p)\boldsymbol{\gamma}(\mathrm{d}p|D,c)\sigma(\mathrm{d}D|c')$ is bounded for all $c' \in C$, by the envelope theorem (Milgrom and Segal, 2002), for any $c \in C$

$$U(c) = U(\bar{c}) - \int_c^{\bar{c}} u_1(z,z)\,\mathrm{d}z = U(\bar{c}) + \int_c^{\bar{c}} \left(\int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} D(p)\boldsymbol{\gamma}(\mathrm{d}p|D,c) \right) \sigma(\mathrm{d}D|z) \right) \mathrm{d}z.$$

Also, by definition,

$$U(c) = \int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} (p-c)D(p)\boldsymbol{\gamma}(\mathrm{d}p|D,c) \right) \sigma(\mathrm{d}D|c) - \tau(c).$$

Rearranging, and letting $\bar{\tau} := U(\bar{c})$, for any $c \in C$,

$$\tau(c) = \int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} (p-c)D(p)\boldsymbol{\gamma}(\mathrm{d}p|D,c) \right) \sigma(\mathrm{d}D|c) - \int_c^{\bar{c}} \left(\int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} D(p)\boldsymbol{\gamma}(\mathrm{d}p|D,z) \right) \sigma(\mathrm{d}D|z) \right) \mathrm{d}z - \bar{\tau},$$

which proves assertion 1. Furthermore, since $u(c,c')$ is affine in c for all c', U is convex as it is a pointwise maximum of a family of affine functions. Therefore, it's (almost everywhere) derivative

$$-\int_{\mathcal{D}} \int_{\mathbb{R}_+} D(p)\boldsymbol{\gamma}(\mathrm{d}p|D,c)\sigma(\mathrm{d}D|c)$$

is nondecreasing. This proves assertion 2.

Conversely, given a mechanism $(\sigma, \tau, \boldsymbol{\gamma})$ that satisfies assertions 1 and 2, for any $c, c' \in C$,

$$\begin{aligned}
& u(c,c) - u(c,c') \\
=& \left(\int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} (p-c)D(p)\boldsymbol{\gamma}(\mathrm{d}p|D,c) \right) \sigma(\mathrm{d}D|c) - \tau(c) \right) \\
& - \left(\int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} (p-c)D(p)\boldsymbol{\gamma}(\mathrm{d}p|D,c') \right) \sigma(\mathrm{d}D|c') - \tau(c') \right) \\
=& \int_c^{c'} \int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} D(p)\boldsymbol{\gamma}(\mathrm{d}p|D,z) \right) \sigma(\mathrm{d}D|z)\,\mathrm{d}z - \int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} D(p)\boldsymbol{\gamma}(\mathrm{d}p|D,c') \right) \sigma(\mathrm{d}D|c')(c'-c) \\
=& \int_c^{c'} \left(\int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} D(p)\boldsymbol{\gamma}(\mathrm{d}p|D,z) \right) \sigma(\mathrm{d}D|z) - \int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} D(p)\boldsymbol{\gamma}(\mathrm{d}p|D,c') \right) \sigma(\mathrm{d}D|c') \right) \mathrm{d}z \\
\geq& 0,
\end{aligned}$$

where the inequality follows from assertion 2. As such, the mechanism $(\sigma, \tau, \boldsymbol{\gamma})$ is indeed incentive compatible. ∎

With Lemma 14, the producer's expected profit under an incentive compatible mechanism $(\sigma, \tau, \boldsymbol{\gamma})$ of the price-controlling data broker can be written as

$$U(c) = U(\bar{c}) + \int_c^{\bar{c}} \int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} D(p) \boldsymbol{\gamma}(\mathrm{d}p|D, z) \right) \sigma(\mathrm{d}D|z) \, \mathrm{d}z.$$

As such, an incentive compatible mechanism is individually rational if and only if

$$U(\bar{c}) + \int_c^{\bar{c}} \int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} D(p) \boldsymbol{\gamma}(\mathrm{d}p|D, z) \right) \sigma(\mathrm{d}D|z) \, \mathrm{d}z \geq \bar{\pi} + \int_c^{\bar{c}} D_0(\overline{\boldsymbol{p}}_0(z)) \, \mathrm{d}z.$$

Also, for any incentive compatible mechanism $(\sigma, \tau, \boldsymbol{\gamma})$, the price-controlling data broker's expected revenue can be written as

$$\mathbb{E}[\tau(c)] = \int_C \left(\int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} (p - \phi_G(c)) D(p) \boldsymbol{\gamma}(\mathrm{d}p|D, c) \right) \sigma(\mathrm{d}D|c) \right) G(\mathrm{d}c) - U(\bar{c}).$$

Therefore, the price-controlling data broker's revenue maximization problem can be written as

$$\begin{aligned}
\sup_{\sigma, \boldsymbol{\gamma}} & \int_C \left(\int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} (p - \phi_G(c)) D(p) \boldsymbol{\gamma}(\mathrm{d}p|D, c) \right) \sigma(\mathrm{d}D|c) \right) G(\mathrm{d}c) - \bar{\pi} \\
\text{s.t. } & c \mapsto \int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} D(p) \boldsymbol{\gamma}(\mathrm{d}p|D, c) \right) \sigma(\mathrm{d}D|c) \text{ is nonincreasing,} \\
& \int_c^{\bar{c}} \int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} D(p) \boldsymbol{\gamma}(\mathrm{d}p|D, z) \right) \sigma(\mathrm{d}D|z) \, \mathrm{d}z \geq \int_c^{\bar{c}} D_0(\overline{\boldsymbol{p}}_0(z)) \, \mathrm{d}z, \, \forall c \in C,
\end{aligned}$$

where the supremum is taken over all segmentation schemes $\sigma \in \mathcal{S}^C$ and all measurable function $\boldsymbol{\gamma}$ that maps from C to the collection of transition kernels from $\mathcal{D}$ to $\Delta(\mathbb{R}_+)$.

In what follows, let Γ be the collection of transition kernels that maps from $\mathcal{D}$ to $\Delta(\mathbb{R}_+)$. Let $\bar{s} \in \mathcal{S}$ denote the value-revealing segmentation and let $\bar{\sigma} \in \mathcal{S}^C$ be the segmentation scheme such that $\bar{\sigma}(c) = \bar{s}$ for all $c \in C$. Furthermore, for any $q \in [0, 1]$, let $\rho_q := D_0^{-1}(q)$, where D_0^{-1} is defined by (14). Notice that by definition of D_0^{-1},

$$q \in [D_0(\rho_q^+), D_0(\rho_q)].$$

If $D_0(\rho_q) = D_0(\rho_q^+)$, then let $\tilde{\gamma}^q \in \Delta(\mathbb{R}_+)^V$ be defined as

$$\tilde{\gamma}^q(\cdot|v) := \delta_{\{v\}}, \, \forall v \in V.$$

On the other hand, if $D_0(\rho_q) > D_0(\rho_q^+)$, then define $\tilde{\gamma}^q \in \Delta(\mathbb{R}_+)^V$ as

$$\tilde{\gamma}^q(\cdot|v) := \begin{cases} \delta_{\{v\}}, & \text{if } v \neq \rho_q \\ \frac{q - D_0(\rho_q^+)}{D_0(\rho_q) - D_0(\rho_q^+)} \delta_{\{v\}} + \frac{D_0(\rho_q) - q}{D_0(\rho_q) - D_0(\rho_q^+)} \delta_{\{\bar{v}\}}, & \text{if } v = \rho_q \end{cases}, \, \forall v \in V.$$

Finally, let $\gamma^q \in \Gamma$ be defined as

$$\gamma^q(A|D) := \int_V \tilde{\gamma}^q(A|v) D(\mathrm{d}v),$$

for any measurable $A \subseteq V$ and for any $D \in \mathcal{D}$. In other words, combining the segmentation $\bar{s}$ and the randomized price γ^q, this means that when the producer uses the randomized price γ^q under segmentation

$\bar{s}$, then all the consumers with values above the $(1-q)$th-percentile buy the product by paying exactly their values while the other consumers do not buy, so that the traded quantity is exactly q. That is,

$$\int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} D(p)\gamma^q(\mathrm{d}p|D) \right) \bar{s}(\mathrm{d}D) = q. \tag{25}$$

With this notation, I now introduce the second auxiliary lemma.

Lemma 15. *For any $q \in [0,1]$, let $\bar{R}(q)$ be the value of the maximization problem*

$$\begin{aligned} \sup_{s\in\mathcal{S},\gamma\in\Gamma} &\int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} pD(p)\boldsymbol{\gamma}(\mathrm{d}p) \right) s(\mathrm{d}D) \\ \text{s.t.} &\int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} D(p)\boldsymbol{\gamma}(\mathrm{d}p) \right) s(\mathrm{d}D) \leq q. \end{aligned} \tag{26}$$

Then

$$\bar{R}(q) = \int_0^q D_0^{-1}(y)\,\mathrm{d}y,$$

where D_0^{-1} is defined by (14). Moreover, $(\bar{s}, \gamma^q)$ is a solution of (26).

Proof. Consider the dual problem of (26). That is, for any $\nu \geq 0$, let

$$\begin{aligned} d(\nu) := &\sup_{s\in\mathcal{S},\gamma\in\Gamma} \left[\int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} pD(p)\gamma(\mathrm{d}p|D) \right) s(\mathrm{d}D) + \nu \left(q - \int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} pD(p)\gamma(\mathrm{d}p|D) \right) s(\mathrm{d}D) \right) \right] \\ = &\sup_{s\in\mathcal{S},\gamma\in\Gamma} \int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} (p-\nu)D(p)\gamma(\mathrm{d}p|D) \right) s(\mathrm{d}D) + \nu q. \end{aligned}$$

Clearly, $d(\nu) \geq \bar{R}(q)$ for any $\nu \geq 0$. Thus, by weak duality, to solve (26), it suffices to find ν^* and (s^*, γ^*) such that (s^*, γ^*) is feasible in the primal problem (26), (s^*, γ^*) solves the dual problem

$$\sup_{s\in\mathcal{S},\gamma\in\Gamma} \int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} (p-\nu^*)D(p)\gamma(\mathrm{d}p|D) \right) s(\mathrm{d}D) \tag{27}$$

and that the complementary slackness condition

$$\nu^* \left[q - \int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} pD(p)\gamma^*(\mathrm{d}p|D) \right) s^*(\mathrm{d}D) \right] = 0 \tag{28}$$

holds. Since this would imply that

$$\bar{R}(q) \leq d^* = \inf_{\lambda\geq 0} d(\lambda) \leq d^*(\nu^*) = \bar{R}(q)$$

and hence (s^*, γ^*) must be a solution to (26).

To this end, let

$$\nu^* := D_0^{-1}(q).$$

and consider the pair $(\bar{s}, \gamma^q)$. Notice that by definition, $(\bar{s}, \gamma^q)$ perfectly price-discriminates all the consumers with $v > \nu^*$ and does not sell to any consumers with $v < \nu^*$. Therefore,

$$\int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} (p-\nu^*)D(p)\gamma(\mathrm{d}p|D) \right) \bar{s}(\mathrm{d}D) = \int_V (v-\nu^*)^+ D_0(\mathrm{d}v)$$

Furthermore, notice that for any $s \in \mathcal{S}$ and any $\gamma \in \Gamma$

$$\begin{aligned}
&\int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} (p - \nu^*) D(p) \gamma(\mathrm{d}p|D) \right) s(\mathrm{d}D) \\
\leq &\int_{\mathcal{D}} \max_{p \in \mathbb{R}_+} (p - \nu^*) D(p) s(\mathrm{d}D) \\
\leq &\int_{V} (v - \nu^*)^+ D_0(\mathrm{d}v).
\end{aligned}$$

Therefore, $(\bar{s}, \gamma^q)$ solves the dual problem (27). On the other hand, by (25),

$$\int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} D(p) \gamma^q(\mathrm{d}p|D) \right) \bar{s}(\mathrm{d}D) = q.$$

Together with $\bar{s} \in \mathcal{S}$ and $\gamma^q \in \Gamma$, it follows that $(\bar{s}, \gamma^q)$ is feasible in the primal problem (26) and, furthermore, the complementary slackness condition (28) also holds. Thus, $(\bar{s}, \gamma^q)$ is a solution to the primal problem (26).

Finally, notice that by the definition of D_0^{-1} and $(\bar{s}, \gamma^q)$,

$$\int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} pD(p) \gamma^q(\mathrm{d}p|D) \right) \bar{s}(\mathrm{d}D) = \int_0^q D_0^{-1}(y)\, \mathrm{d}y.$$

This completes the proof. ∎

With the two auxiliary lemmas above, the price-controlling data broker's problem can be effectively reduced to a one-dimensional screening problem with type-dependent individual rationality constraints. As stated by Lemma 16 below.

Lemma 16. *There exists an incentive feasible mechanism that maximizes the price-controlling data broker's revenue. Furthermore, the price-controlling data broker's optimal revenue is*

$$\begin{aligned}
R^* = \max_{\boldsymbol{q} \in \mathcal{Q}} &\int_C \left(\int_0^{\boldsymbol{q}(c)} (D_0^{-1}(q) - \phi_G(c))\, \mathrm{d}q \right) G(\mathrm{d}c) - \bar{\pi} \\
\text{s.t. } &\bar{\pi} + \int_c^{\bar{c}} \boldsymbol{q}(z)\, \mathrm{d}z \geq \pi_0(c), \, \forall c \in C,
\end{aligned} \tag{29}$$

where $\mathcal{Q}$ is the collection of nonincreasing functions in $[0,1]^C$.

Proof. Consider any incentive feasible mechanism $(\sigma, \tau, \boldsymbol{\gamma})$ for the price-controlling data broker, I will first show that there exists $\boldsymbol{q} \in [0,1]^C$ such that the mechanism $(\bar{\sigma}, \tau^{\boldsymbol{q}}, \boldsymbol{\gamma}^{\boldsymbol{q}})$ generates weakly higher revenue for price-controlling data broker and is incentive feasible, where

$$\boldsymbol{\gamma}^{\boldsymbol{q}}(c) := \gamma^{\boldsymbol{q}(c)}, \, \forall \, c \in C$$

and $\tau^{\boldsymbol{q}}$ is the transfer determined by $(\bar{\sigma}, \boldsymbol{\gamma}^{\boldsymbol{q}})$ when the constant is chosen so that $U(\bar{c}) = \bar{\pi}$ according to Lemma 14. Then, I will show that maximizing revenue across the family of incentive feasible mechanisms $(\bar{\sigma}, \tau^{\boldsymbol{q}}, \boldsymbol{\gamma}^{\boldsymbol{q}})$ is equivalent to solving (29). Finally, the existence of the optimal mechanism can then be ensured by the existence of the solution of (29).

To this end, for any $c \in C$, let

$$\boldsymbol{q}(c) := \int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} D(p) \boldsymbol{\gamma}(\mathrm{d}p|D,c) \right) \sigma(\mathrm{d}D|c).$$

By Lemma 14, incentive compatibility of $(\sigma, \tau, \boldsymbol{\gamma})$ implies that $\boldsymbol{q} \in [0,1]^C$ is nonincreasing and, by (25), for any $c \in C$,

$$\int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} D(p) \boldsymbol{\gamma}^{\boldsymbol{q}}(\mathrm{d}p|D,c) \right) \bar{\sigma}(\mathrm{d}D|c) = \int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} D(p) \gamma^{\boldsymbol{q}(c)}(\mathrm{d}p|D) \right) \bar{s}(\mathrm{d}D) = \boldsymbol{q}(c).$$

Thus, by Lemma 15, $(\bar{\sigma}(c), \boldsymbol{\gamma}^{\boldsymbol{q}}(c))$ solves the problem (26) with the quantity constraint being $\boldsymbol{q}(c)$ and hence, since $(\sigma(c), \boldsymbol{\gamma}(c))$ is also feasible in this problem,

$$\int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} pD(p) \boldsymbol{\gamma}(\mathrm{d}p|D,c) \right) \sigma(\mathrm{d}D|c) \le \int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} pD(p) \boldsymbol{\gamma}^{\boldsymbol{q}}(\mathrm{d}p|D,c) \right) \bar{\sigma}(\mathrm{d}D|c) = \bar{R}(\boldsymbol{q}(c)). \tag{30}$$

As a result,

$$\begin{aligned}
&\int_C \left(\int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} (p - \phi_G(c)) D(p) \boldsymbol{\gamma}(\mathrm{d}p|D,c) \right) \sigma(\mathrm{d}x|c) \right) G(\mathrm{d}c) \\
&= \int_C \left(\int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} pD(p) \boldsymbol{\gamma}(\mathrm{d}p|D,c) \right) \sigma(\mathrm{d}x|c) \right) G(\mathrm{d}c) - \int_C \phi_G(c) \boldsymbol{q}(c) G(\mathrm{d}c) \\
&\le \int_C (\bar{R}(\boldsymbol{q}(c)) - \phi_G(c) \boldsymbol{q}(c)) G(\mathrm{d}c) \\
&= \int_C \left(\int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} pD(p) \boldsymbol{\gamma}^{\boldsymbol{q}}(\mathrm{d}p|D,c) \right) \bar{\sigma}(\mathrm{d}x|c) \right) G(\mathrm{d}c) - \int_C \phi_G(c) \boldsymbol{q}(c) G(\mathrm{d}c) \\
&= \int_C \left(\int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} (p - \phi_G(c)) D(p) \boldsymbol{\gamma}^{\boldsymbol{q}}(\mathrm{d}p|D,c) \right) \bar{\sigma}(\mathrm{d}x|c) \right) G(\mathrm{d}c),
\end{aligned}$$

where the first and the third equalities follows from the definition of $\boldsymbol{q}(c)$ and from (25), and the inequality and the second equality follows from (30). Moreover, by (25), since $\boldsymbol{q}$ is nonincreasing, the function

$$c \mapsto \int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} D(p) \boldsymbol{\gamma}^{\boldsymbol{q}}(\mathrm{d}p|D,c) \right) \bar{\sigma}(\mathrm{d}D|c)$$

is nonincreasing. Together with Lemma 14 and individual rationality of $(\sigma, \tau, \boldsymbol{\gamma})$, for any $c \in C$,

$$\begin{aligned}
\int_c^{\bar{c}} \left(\int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} D(p) \boldsymbol{\gamma}^{\boldsymbol{q}}(\mathrm{d}p|D,z) \right) \bar{\sigma}(\mathrm{d}D|z) \right) \mathrm{d}z &= \int_c^{c'} \boldsymbol{q}(z) \, \mathrm{d}z \\
&= \int_c^{\bar{c}} \left(\int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} D(p) \boldsymbol{\gamma}(\mathrm{d}p|D,z) \right) \sigma(\mathrm{d}D|z) \right) \mathrm{d}z \\
&\ge \int_c^{\bar{c}} D_0(\overline{\boldsymbol{p}}_0(z)) \, \mathrm{d}z,
\end{aligned}$$

these imply that $(\bar{\sigma}, \tau^q, \boldsymbol{\gamma}^{\boldsymbol{q}})$ is incentive feasible.

Now notice that by (25) and Lemma 15, for any $\boldsymbol{q} \in [0,1]^C$ and for any $c \in C$,

$$\int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} (p - \phi_G(c)) D(p) \boldsymbol{\gamma}^{\boldsymbol{q}}(\mathrm{d}p|D,c) \right) \bar{\sigma}(\mathrm{d}D|c) = \bar{R}(\boldsymbol{q}(c)) = \int_0^{\boldsymbol{q}(c)} D_0^{-1}(q) \, \mathrm{d}q.$$

On the other hand, by (25) and by Lemma 15, $(\bar{\sigma}, \tau^{\boldsymbol{q}}, \boldsymbol{\gamma}^{\boldsymbol{q}})$ is incentive feasible if and only if $\boldsymbol{q}$ is nonincreasing and

$$\int_c^{\bar{c}} \boldsymbol{q}(z)\,\mathrm{d}z \geq \int_c^{\bar{c}} D_0(\overline{\boldsymbol{p}}_0(z))\,\mathrm{d}z,\ \forall c \in C.$$

Therefore, maximizing revenue among all incentive feasible mechanism is equivalent to solving (29).

Finally, notice that for the maximization problem (29), endow the set of nonincreasing functions with the L^1 norm. Helly's selection theorem and the Lebesgue dominated convergence theorem then imply that this set is compact. Furthermore, since for any sequence $\{q_n\}$ that converges to q in the L^1 norm, there exists a subsequence $\{q_{n_k}\}$ that converges to q pointwise, by the Lebesgue dominated convergence again, the objective function of (29) is continuous and the feasible set is a closed subset of a compact set, and hence is itself compact. Together, the problem (29) has a solution. This completes the proof. ■

With Lemma 16, the price-controlling data broker's revenue maximization problem can be solved explicitly.

Proof of Proposition 2. Recall that $\mathcal{Q} \subset [0,1]^C$ denotes the set of nonincreasing functions $[0,1]^C$. Using (15) and Lemma 16, rewrite the price-controlling data broker's problem (29) as

$$\begin{aligned} &\sup_{\boldsymbol{q}\in\mathcal{Q}} \int_C \left(\int_0^{\boldsymbol{q}(c)} (D_0^{-1}(q) - \phi_G(c))\,\mathrm{d}q\right) G(\mathrm{d}c) - \bar{\pi} \\ &\text{s.t. } \bar{\pi} + \int_c^{\bar{c}} \boldsymbol{q}(z)\,\mathrm{d}z \geq \bar{\pi} + \int_c^{\bar{c}} D_0(\boldsymbol{p}_0(z))\,\mathrm{d}z. \end{aligned} \tag{31}$$

Let R^* be the value of (31) and write objective function of (31) as

$$R(\boldsymbol{q}) := \int_C \left(\int_0^{\boldsymbol{q}(c)} (D_0^{-1}(q) - \phi_G(c))\,\mathrm{d}q\right) G(\mathrm{d}c) - \bar{\pi},\ \forall q \in \mathcal{Q}.$$

Consider the dual problem of (31). That is, for any Borel measure μ on C, let

$$d(\mu) := \sup_{\boldsymbol{q}\in\mathcal{Q}} \left[R(\boldsymbol{q}) + \int_C \left(\int_c^{\bar{c}} (\boldsymbol{q}(z) - D_0(\overline{\boldsymbol{p}}_0(z)))\,\mathrm{d}z\right) \mu(\mathrm{d}c)\right]$$

and let

$$d^* := \inf_{\mu} d(\mu),$$

where the infimum is taken over all Borel measures on C. Then clearly

$$d^* \geq R^*.$$

Moreover, if there exists a Borel measure μ^* on C and a feasible choice $\boldsymbol{q}^* \in \mathcal{Q}$ of the primal problem (31) such that $d(\mu^*) = R(\boldsymbol{q}^*)$, then

$$R^* \leq d^* \leq d(\mu^*) = R(\boldsymbol{q}^*) \leq R^*,$$

and hence $\boldsymbol{q}^* \in \mathcal{Q}$ is a solution of the primal problem (31). As a result, to solve (31), it suffices to find a Borel measure μ^* and a feasible $\boldsymbol{q}^* \in \mathcal{Q}$ such that $\boldsymbol{q}^*$ is a solution of

$$\sup_{\boldsymbol{q}\in\mathcal{Q}} \left[R(\boldsymbol{q}) + \int_C \left(\int_c^{\bar{c}} (\boldsymbol{q}(z) - D_0(\overline{\boldsymbol{p}}_0(z)))\,\mathrm{d}z\right) \mu^*(\mathrm{d}c)\right] \tag{32}$$

and that

$$\int_C \left(\int_c^{\overline{c}} (\boldsymbol{q}^*(z) - D_0(\overline{\boldsymbol{p}}_0(z)))\, \mathrm{d}z \right) \mu^*(\mathrm{d}c) = 0. \tag{33}$$

To this end, define $M^* \in [0,1]^C$ as the following

$$M^*(c) := \lim_{z \downarrow c} g(z)(\phi_G(z) - \overline{\boldsymbol{p}}_0(z))^+, \, \forall c \in C. \tag{34}$$

By definition, M^* is right-continuous. Also, by Assumption 1, M^* is nondecreasing and hence M^* a CDF. Let μ^* be the Borel measure induced by M^*. Notice that $\mathrm{supp}(\mu^*) = [c^*, \overline{c}]$, where

$$c^* := \inf\{c \in C : \phi_G(c) > \overline{\boldsymbol{p}}_0(c)\}.$$

Notice that for any $q \in \mathcal{Q}$, by interchanging the order of integrals,

$$\int_C \left(\int_c^{\overline{c}} (\boldsymbol{q}(z) - D_0(\overline{\boldsymbol{p}}_0(z)))\, \mathrm{d}z \right) \mu^*(\mathrm{d}c) = \int_C M^*(c)(\boldsymbol{q}(c) - D_0(\overline{\boldsymbol{p}}_0(c)))\, \mathrm{d}c$$

and therefore for any $\boldsymbol{q} \in \mathcal{Q}$,

$$\begin{aligned}
&R(\boldsymbol{q}) + \int_C \left(\int_c^{\overline{c}} (\boldsymbol{q}(z) - D_0(\overline{\boldsymbol{p}}_0(z)))\, \mathrm{d}z \right) \mu^*(\mathrm{d}c) \\
=& \int_C \left(\int_0^{\boldsymbol{q}(c)} (D_0^{-1}(q) - \phi_G(c))\, \mathrm{d}q \right) G(\mathrm{d}c) - \bar{\pi} + \int_C M^*(c)(\boldsymbol{q}(c) - D_0(\overline{\boldsymbol{p}}_0(c)))\, \mathrm{d}c \\
=& \int_C \left(\int_0^{\boldsymbol{q}(c)} (D_0^{-1}(q) - \overline{\phi}_G(c))\, \mathrm{d}q \right) G(\mathrm{d}c) - \bar{\pi} - \int_C M^*(c) D_0(\overline{\boldsymbol{p}}_0(c))\, \mathrm{d}c,
\end{aligned}$$

where $\overline{\phi}_G := \min\{\phi_G, \overline{\boldsymbol{p}}_0\}$. As only the first term depends on the choice variable $\boldsymbol{q}$, the solution of (32) is the same as the solution of

$$\sup_{\boldsymbol{q} \in \mathcal{Q}} \int_C \left(\int_0^{\boldsymbol{q}(c)} (D_0^{-1}(q) - \overline{\phi}_G(c))\, \mathrm{d}q \right) G(\mathrm{d}c). \tag{35}$$

To solve (35), consider first the case when G is regular so that ϕ_G is nondecreasing. In this case, notice that for any $\boldsymbol{q} \in \mathcal{Q}$ and for all $c \in C$,

$$\int_0^{\boldsymbol{q}(c)} (D_0^{-1}(q) - \overline{\phi}_G(c))\, \mathrm{d}q \leq \int_0^{\boldsymbol{q}^{\overline{\phi}_G}(c)} (D_0^{-1}(q) - \overline{\phi}_G(c))\, \mathrm{d}q,$$

where for any function ψ,

$$\boldsymbol{q}^{\psi}(c) := \sup\{y \in [0,1] : D_0^{-1}(q) \geq \psi(c)\}.$$

Moreover, the function $\overline{\phi}_G$ is nondecreasing since both ϕ_G and $\overline{\boldsymbol{p}}_0$ are nondecreasing, also, the function D_0^{-1} is nonincreasing. As a result, $\boldsymbol{q}^{\overline{\phi}_G} \in \mathcal{Q}$ is a solution of (35) and thus is a solution of (32). Furthermore, by definition, for any $c \in C$,

$$D_0(\overline{\phi}_G(c)) = \boldsymbol{q}^{\overline{\phi}_G}(c)$$

In particular, since $\overline{\phi}_G \leq \overline{\boldsymbol{p}}_0$, $\boldsymbol{q}^{\overline{\phi}_G}$ is feasible in the primal problem (31). That is, for any $c \in C$,

$$\begin{aligned}
\bar{\pi} + \int_c^{\overline{c}} \boldsymbol{q}^{\overline{\phi}_G}(z)\, \mathrm{d}z \geq& \bar{\pi} + \int_c^{\overline{c}} D_0(\overline{\phi}_G(z))\, \mathrm{d}z \\
\geq& \bar{\pi} + \int_c^{\overline{c}} D_0(\overline{\boldsymbol{p}}_0(z))\, \mathrm{d}z.
\end{aligned}$$

Finally, notice that since $\overline{\phi}_G(c) = \overline{\boldsymbol{p}}_0(c)$ for all $c \in (c^*, \overline{c}]$ and since $M^*(c) = 0$ for all $c \in [\underline{c}, c^*)$, the complementary slackness condition (33) also holds for $\overline{\phi}_G$ and μ^*. That is,

$$\begin{aligned}\int_C M^*(c)(\boldsymbol{q}^{\overline{\phi}_G}(c) - D_0(\overline{\boldsymbol{p}}_0(c)))\,\mathrm{d}c =& \int_{c^*}^{\overline{c}} M^*(c)(D_0(\overline{\phi}_G(c)) - D_0(\overline{\boldsymbol{p}}_0(c)))\,\mathrm{d}c\\ =& \int_{c^*}^{\overline{c}} M^*(c)(D_0(\overline{\boldsymbol{p}}_0(c)) - D_0(\overline{\boldsymbol{p}}_0(c)))\,\mathrm{d}c\\ =&0.\end{aligned}$$

Together, when G is regular, $\boldsymbol{q}^{\overline{\phi}_G} \equiv D_0 \circ \overline{\phi}_G$ solves the primal problem (31).

Now consider the case for a general G, to solve (35) Let φ_G be the ironed virtual valuation. That is, φ_G is defined by the following procedure: Let $h : [0,1] \to \mathbb{R}_+$ be defined as

$$h(q) := \phi_G(G^{-1}(q)) = G^{-1}(q) + \frac{q}{g(G^{-1}(q))}.$$

and define $H : [0,1] \to \mathbb{R}_+$, $K : [0,1] \to \mathbb{R}_+$ as

$$H(q) := \int_0^q h(s)\,\mathrm{d}s$$

and

$$K := \mathrm{co}(H).$$

Finally, for every $q \in [0,1]$ let $k(q) := K'(q)$. φ_G is then defined as

$$\varphi_G(c) := k(G(c)).$$

Also, let $\overline{\varphi}_G := \min\{\varphi_G, \overline{\boldsymbol{p}}_0\}$. With this definition, notice that for any $\boldsymbol{q} \in \mathcal{Q}$,

$$\begin{aligned}&\int_C \left(\int_0^{\boldsymbol{q}(c)} (D_0^{-1}(q) - \overline{\phi}_G(c))\,\mathrm{d}q\right) G(\mathrm{d}c)\\ =&\int_C \left(\int_0^{\boldsymbol{q}(c)} (D_0^{-1}(q) - \overline{\varphi}_G(c))\,\mathrm{d}q\right) G(\mathrm{d}c) + \int_C (\overline{\varphi}_G(c) - \overline{\phi}_G(c))\boldsymbol{q}(c)G(\mathrm{d}c).\end{aligned} \tag{36}$$

Moreover, using integration by parts, since $K(0) = H(0)$ and $K(1) = H(1)$,

$$\begin{aligned}\int_C (\overline{\varphi}_G(c) - \overline{\phi}_G(c))\boldsymbol{q}(c)G(\mathrm{d}c) =& \int_C (\varphi_G(c) - \phi_G(c))\boldsymbol{q}(c)G(\mathrm{d}c)\\ =& -\int_C (K(G(c)) - H(G(c)))\boldsymbol{q}(\mathrm{d}c)\\ \leq& 0,\end{aligned} \tag{37}$$

where the first equality follows from the observation that $\overline{\phi}_G(c) = \overline{\varphi}_G(c) = \phi_G(c) = \varphi_G(c) = \overline{\boldsymbol{p}}_0(c)$ for all $c \geq c^*$, which is due to Assumption 1, and the inequality follows from the fact that $K = \mathrm{co}(H)$ and that $\boldsymbol{q}$ is nonincreasing for any $\boldsymbol{q} \in \mathcal{Q}$.

On the other hand, notice that for any $\boldsymbol{q} \in \mathcal{Q}$ and for all $c \in C$,

$$\int_0^{\boldsymbol{q}(c)} (D_0^{-1}(q) - \overline{\varphi}_G(c))\,\mathrm{d}q \leq \int_0^{\boldsymbol{q}^{\overline{\varphi}_G}(c)} (D_0^{-1}(q) - \overline{\varphi}_G(c))\,\mathrm{d}q,$$

and hence

$$\int_C \left(\int_0^{\boldsymbol{q}(c)} (D_0^{-1}(q) - \overline{\varphi}_G(c))\,\mathrm{d}q\right) G(\mathrm{d}c) \leq \int_C \left(\int_0^{\boldsymbol{q}^{\overline{\varphi}_G(c)}} (D_0^{-1}(q) - \overline{\varphi}_G(c))\,\mathrm{d}q\right), \forall \boldsymbol{q} \in \mathcal{Q}.$$

In addition, since $\overline{\varphi}_G(c) = \overline{\phi}_G(c) = \overline{\boldsymbol{p}}_0(c)$ for all $c \in (c^*, \overline{c}]$ and since $K(G(c)) < H(G(c))$ on an interval $[c_1, c_2]$ if and only if $\overline{\varphi}_G$ is a constant on that interval, which implies that $\boldsymbol{q}^{\overline{\varphi}_G}$ is a constant on that interval, it must be that

$$\int_C (\overline{\varphi}_G(c) - \overline{\phi}_G(c))\boldsymbol{q}^{\overline{\varphi}_G}(c) G(\mathrm{d}c) = -\int_C (K(G(c)) - H(G(c)))\boldsymbol{q}^{\overline{\varphi}_G}(\mathrm{d}c) = 0. \tag{38}$$

Together with (36) and (37), for any $\boldsymbol{q} \in \mathcal{Q}$,

$$\int_C \left(\int_0^{\boldsymbol{q}(c)} (D_0^{-1}(q) - \overline{\phi}_G(c))\,\mathrm{d}q\right) G(\mathrm{d}c) \leq \int_C \left(\int_0^{\boldsymbol{q}^{\overline{\varphi}_G}(c)} (D_0^{-1}(q) - \overline{\phi}_G(c))\,\mathrm{d}q\right) G(\mathrm{d}c).$$

Also, since $\overline{\varphi}_G$ is nondecreasing by definition, $\boldsymbol{q}^{\overline{\varphi}_G}$ is indeed a solution of (35) and hence a solution of (32).

Moreover, by definition, for any $c \in C$,

$$\boldsymbol{q}^{\overline{\varphi}_G}(c) = D_0(\overline{\varphi}_G(c)). \tag{39}$$

Thus, since $\overline{\varphi}_G \leq \overline{\boldsymbol{p}}_0$,

$$\begin{aligned}\bar{\pi} + \int_c^{\overline{c}} \boldsymbol{q}^{\overline{\varphi}_G}(z)\,\mathrm{d}z =& \bar{\pi} + \int_c^{\overline{c}} D_0(\overline{\varphi}_G(c))\,\mathrm{d}z \\ \geq& \bar{\pi} + \int_c^{\overline{c}} D_0(\overline{\boldsymbol{p}}_0(z))\,\mathrm{d}z, \forall c \in C\end{aligned}$$

and hence $\boldsymbol{q}^{\overline{\varphi}_G} \in \mathcal{Q}$ is feasible choice in the primal problem (31). Furthermore, since $M^*(c) = 0$ for all $c \in [\underline{c}, c^*)$ and since $\overline{\varphi}_G(c) = \overline{\boldsymbol{p}}_0(c)$ for all $c \in (c^*, \overline{c}]$, the complementary slackness condition (33) is also satisfied. That is,

$$\begin{aligned}\int_C M^*(c)(\boldsymbol{q}^{\overline{\varphi}_G}(c) - D_0(\overline{\boldsymbol{p}}_0(c)))\,\mathrm{d}c =& \int_C M^*(c)(D_0(\overline{\varphi}_G(c)) - D_0(\overline{\boldsymbol{p}}_0(c)))\,\mathrm{d}c \\ =& \int_{c^*}^{\overline{c}} M^*(c)(D_0(\overline{\boldsymbol{p}}_0(c)) - D_0(\overline{\boldsymbol{p}}_0(c)))\,\mathrm{d}c \\ =& 0.\end{aligned}$$

Together, $\boldsymbol{q}^{\overline{\varphi}_G} \equiv D_0 \circ \overline{\varphi}_G$ is indeed a solution of (31).

Finally, by (39), it then follows that

$$\begin{aligned}R^* =& \int_C \left(\int_0^{\boldsymbol{q}^{\overline{\varphi}_G}(c)} (D_0^{-1}(q) - \phi_G(c))\,\mathrm{d}q\right) G(\mathrm{d}c) - \bar{\pi} \\ =& \int_C \left(\int_0^{D_0(\overline{\varphi}_G(c))} (D_0^{-1}(q) - \phi_G(c))\,\mathrm{d}q\right) G(\mathrm{d}c) - \bar{\pi} \\ =& \int_C \left(\int_{\{v \geq \overline{\varphi}_G(c)\}} (v - \phi_G(c)) D_0(\mathrm{d}v)\right) G(\mathrm{d}c) - \bar{\pi}.\end{aligned}$$

The see that any solution of the price-controlling data broker's problem must induce $\overline{\varphi}_G(c)$-quasi-perfect price discrimination for G almost all $c \in C$, consider any optimal mechanism $(\sigma, \tau, \boldsymbol{\gamma})$ of the price-controlling data broker. By optimality, it must be that $\mathbb{E}_G[\tau(c)] = R^*$ and that the indirect utility of the producer with marginal cost $\bar{c}$ is $\bar{\pi}$. Also, by Lemma 14, it must be that

$$\begin{aligned}R^* =& \mathbb{E}_G[\tau(c)]\\ =& \int_C \left(\int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} (p - \phi_G(c)) D(p) \boldsymbol{\gamma}(\mathrm{d}p|D, c) \right) \sigma(\mathrm{d}D|c) \right) G(\mathrm{d}c) - \bar{\pi},\end{aligned}$$

which implies that

$$\int_C \left(\int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} (p - \phi_G(c)) D(p) \boldsymbol{\gamma}(\mathrm{d}p|D, c) \right) \sigma(\mathrm{d}D|c) \right) G(\mathrm{d}c) = \int_C \left(\int_{\{v \geq \overline{\varphi}_G(c)\}} (v - \phi_G(c)) D_0(\mathrm{d}v) \right) G(\mathrm{d}c). \tag{40}$$

Thus,

$$\begin{aligned}&\int_C \left(\int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} (p - \overline{\varphi}_G(c)) D(p) \boldsymbol{\gamma}(\mathrm{d}p|D, c) \right) \sigma(\mathrm{d}D|c) \right) G(\mathrm{d}c)\\ &+ \int_C (\overline{\varphi}_G(c) - \phi_G(c)) \left(\int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} D(p) \boldsymbol{\gamma}(\mathrm{d}p|D, c) \right) \sigma(\mathrm{d}D|c) \right) G(\mathrm{d}c)\\ =& \int_C \left(\int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} (p - \phi_G(c)) D(p) \boldsymbol{\gamma}(\mathrm{d}p|D, c) \right) \sigma(\mathrm{d}D|c) \right) G(\mathrm{d}c)\\ =& \int_C \left(\int_{\{v \geq \overline{\varphi}_G(c)\}} (v - \phi_G(c)) D_0(\mathrm{d}v) \right) G(\mathrm{d}c)\\ =& \int_C \left(\int_{\{v \geq \overline{\varphi}_G(c)\}} (v - \overline{\varphi}_G(c)) D_0(\mathrm{d}v) \right) G(\mathrm{d}c) + \int_C (\overline{\varphi}_G(c) - \phi_G(c)) D_0(\overline{\varphi}_G(c)) G(\mathrm{d}c)\end{aligned} \tag{41}$$

where the second equality follows from (40). Moreover, since for any $c \in C$,

$$\begin{aligned}&\int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} (p - \overline{\varphi}_G(c)) D(p) \boldsymbol{\gamma}(\mathrm{d}p|D, c) \right) \sigma(\mathrm{d}D|c)\\ \leq& \int_{\mathcal{D}} \max_{p \in \mathbb{R}_+} [(p - \overline{\varphi}_G(c)) D(p)] \sigma(\mathrm{d}D|c)\\ \leq& \int_V (v - \overline{\varphi}_G(c))^+ D_0(\mathrm{d}v),\end{aligned} \tag{42}$$

it must be that

$$\int_C (\overline{\varphi}_G(c) - \phi_G(c)) \left(\int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} D(p) \boldsymbol{\gamma}(\mathrm{d}p|D, c) \right) \sigma(\mathrm{d}D|c) \right) G(\mathrm{d}c) \geq \int_C (\overline{\varphi}_G(c) - \phi_G(c)) D_0(\overline{\varphi}_G(c)) G(\mathrm{d}c).$$

Together with (37) and (38), we have

$$\begin{aligned}
&\int_C (\overline{\phi}_G(c) - \phi_G(c)) \left(\int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} D(p) \boldsymbol{\gamma}(\mathrm{d}p|D,c) \right) \sigma(\mathrm{d}D|c) \right) G(\mathrm{d}c) \\
\geq &\int_C (\overline{\varphi}_G(c) - \phi_G(c)) \left(\int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} D(p) \boldsymbol{\gamma}(\mathrm{d}p|D,c) \right) \sigma(\mathrm{d}D|c) \right) G(\mathrm{d}c) \\
\geq &\int_C (\overline{\varphi}_G(c) - \phi_G(c)) D_0(\overline{\varphi}_G(c)) G(\mathrm{d}c) \\
= &\int_C (\overline{\phi}_G(c) - \phi_G(c)) D_0(\overline{\varphi}_G(c)) G(\mathrm{d}c).
\end{aligned} \tag{43}$$

Furthermore, since $\overline{\phi}_G(c) = \overline{\boldsymbol{p}}_0(c) \leq \phi_G(c)$ for all $c \in (c^*, \overline{c}]$ and $\overline{\phi}_G(c) = \phi_G(c)$, by the definition of M^* given by (34), and by using integration by parts, (43) is equivalent to

$$\int_C \left(\int_c^{\overline{c}} \left(\int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} D(p) \boldsymbol{\gamma}(\mathrm{d}p|D,z) \right) \sigma(\mathrm{d}D|z) - D_0(\overline{\boldsymbol{p}}_0(z) \right) \mathrm{d}z \right) M^*(\mathrm{d}c) \leq 0$$

Furthermore, since $(\sigma, \tau, \boldsymbol{\gamma})$ is individually rational, for any $c \in C$,

$$\int_c^{\overline{c}} \left(\int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} D(p) \boldsymbol{\gamma}(\mathrm{d}p|D,z) \right) \sigma(\mathrm{d}D|z) \right) \mathrm{d}z \geq \int_c^{\overline{c}} D_0(\overline{\boldsymbol{p}}_0(z)) \, \mathrm{d}z.$$

Thus,

$$\int_C \left(\int_c^{\overline{c}} \left(\int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} D(p) \boldsymbol{\gamma}(\mathrm{d}p|D,z) \right) \sigma(\mathrm{d}D|z) - D_0(\overline{\boldsymbol{p}}_0(z)) \right) \mathrm{d}z \right) M^*(\mathrm{d}c) = 0$$

and hence

$$\int_C (\overline{\varphi}_G(c) - \phi_G(c)) \left(\int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} D(p) \boldsymbol{\gamma}(\mathrm{d}p|D,c) \right) \sigma(\mathrm{d}D|c) \right) G(\mathrm{d}c) = \int_C (\overline{\varphi}_G(c) - \phi_G(c)) D_0(\overline{\varphi}_G(c)) G(\mathrm{d}c),$$

which, together with (41), implies that

$$\begin{aligned}
&\int_C \left(\int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} (p - \overline{\varphi}_G(c)) D(p) \boldsymbol{\gamma}(\mathrm{d}p|D,c) \right) \sigma(\mathrm{d}D|c) \right) G(\mathrm{d}c) \\
= &\int_C \left(\int_{\{v \geq \overline{\varphi}_G(c)\}} (v - \overline{\varphi}_G(c)) D_0(\mathrm{d}v) \right) G(\mathrm{d}c).
\end{aligned}$$

Moreover, by (42), it then follows that for G-almost all $c \in C$,

$$\int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} (p - \overline{\varphi}_G(c)) D(p) \boldsymbol{\gamma}(\mathrm{d}p|D,c) \right) \sigma(\mathrm{d}D|c) = \int_V (v - \overline{\varphi}_G(c))^+ D_0(\mathrm{d}v), \tag{44}$$

which implies that for G-almost all $c \in C$, $(\sigma(c), \boldsymbol{\gamma}(c))$ must induce perfect price discrimination for the economy where the producer's marginal cost is $\overline{\varphi}_G(c)$, or equivalently, $(\sigma(c), \boldsymbol{\gamma}(c))$ must induce $\overline{\varphi}_G(c)$-quasi-perfect price discrimination for G-almost all $c \in C$. This completes the proof. ■

D.2 Proof of Lemma 1

Proof of Lemma 1. For necessity, consider any incentive compatible mechanism (σ, τ). First notice that, by Lemma 5, $\pi_D : C \to \mathbb{R}_+$ is convex and continuous on C for any $D \in \mathcal{D}$ and

$$\pi'_D(c) = -D(\boldsymbol{p}_D(c))$$

for any $\boldsymbol{p} \in \boldsymbol{P}$ and for almost all $c \in C$. Moreover, since for any $D \in \mathcal{D}$ and for any $\boldsymbol{p} \in \boldsymbol{P}$

$$\left|\pi'_D(c)\right| = \left|D(\boldsymbol{p}_D(c))\right| \leq 1,$$

for almost all $c \in C$, the order of integral and differential can be interchanged. That is, for any $c, c' \in C$,

$$\frac{\mathrm{d}}{\mathrm{d}c} \int_{\mathcal{D}} \pi_D(c)\sigma(\mathrm{d}D|c') = \int_{\mathcal{D}} \pi'_D(c)\sigma(\mathrm{d}D|c') = -\int_{\mathcal{D}} D(\boldsymbol{p}_D(c))\sigma(\mathrm{d}D|c'). \tag{45}$$

As such, if Π is defined as

$$\Pi(c, c') := \int_{\mathcal{D}} \pi_D(c)\sigma(\mathrm{d}D|c'),\ \forall c, c' \in C,$$

then $\Pi(\cdot, c')$ is convex for all $c' \in C$. Moreover, (45) implies that

$$\Pi_1(c, c') = -\int_{\mathcal{D}} D(\boldsymbol{p}_D(c))\sigma(\mathrm{d}D|c'), \tag{46}$$

for any $c' \in C$, any $\boldsymbol{p} \in \boldsymbol{P}$, and for almost all $c \in C$. Now let $u(c, c') := \Pi(c, c') - t(c')$ for all $c, c' \in C$ be a producer's interim expected profit if her report is c' and marginal cost is c. By the Lebesgue dominated convergence theorem, $u(\cdot, c')$ is convex and absolutely continuous on C for all $c' \in C$ as π_D is absolutely continuous for all $D \in \mathcal{D}$. Furthermore, since the mechanism (σ, τ) is incentive compatible,

$$U(c) := u(c, c) \geq u(c, c'),\ \forall c, c' \in C.$$

By the envelope theorem again,

$$U(c) = U(\overline{c}) - \int_c^{\overline{c}} u_1(z, z)\,\mathrm{d}z.$$

Moreover, since for any $\boldsymbol{p} \in \boldsymbol{P}$ and for almost all $c \in C$,

$$u_1(c, c) = \Pi_1(c, c) = -\int_{\mathcal{D}} D(\boldsymbol{p}_D(c))\sigma(\mathrm{d}D|c), \tag{47}$$

it must be that for any $\boldsymbol{p} \in \boldsymbol{P}$ and for any $c \in C$,

$$\Pi(c, c) - \tau(c) = U(c) = U(\overline{c}) + \int_c^{\overline{c}} \int_{\mathcal{D}} D(\boldsymbol{p}_D(z))\sigma(\mathrm{d}D|z)\,\mathrm{d}z.$$

Rearranging, and use the definition of Π, it follows that

$$\tau(c) = \int_{\mathcal{D}} \pi_D(c)\sigma(\mathrm{d}D|c) - \int_c^{\overline{c}} \int_{\mathcal{D}} D(\boldsymbol{p}_D(z))\sigma(\mathrm{d}D|z) - U(\overline{c}),$$

which proves assertion 1. Furthermore, by incentive compatibility, for any $c, c' \in C$,

$$\begin{aligned}
0 \leq & U(c) - u(c, c') \\
= & U(c) - (\Pi(c, c') - \tau(c')) \\
= & U(c) - (\Pi(c, c') - \Pi(c', c')) - U(c') \\
= & (U(c) - U(c')) + (\Pi(c', c') - \Pi(c, c')) \\
= & \int_c^{c'} (u_1(z, z) + \Pi_1(z, c')) \, \mathrm{d}z \\
= & \int_c^{c'} \left(\int_{\mathcal{D}} D(\boldsymbol{p}_D(z)) \sigma(\mathrm{d}D|z) - \int_{\mathcal{D}} D(\boldsymbol{p}_D(z)) \sigma(\mathrm{d}D|c') \right) \mathrm{d}z \\
= & \int_c^{c'} \left(\int_{\mathcal{D}} D(\boldsymbol{p}_D(z)) (\sigma(\mathrm{d}D|z) - \sigma(\mathrm{d}D|c')) \right) \mathrm{d}z,
\end{aligned}$$

for any $\boldsymbol{p} \in \boldsymbol{P}$, where the forth equality follows from the fundamental theorem of calculus, and the last equality follows from (46) and (47). This proves assertion 2.

Conversely, suppose that a mechanism (σ, τ) satisfies assertions 1 and 2. Then, for any $c, c' \in C$,

$$\begin{aligned}
& (\Pi(c, c) - \tau(c)) - (\Pi(c, c') - \tau(c')) \\
= & \Pi(c, c) - \tau(c) - (\Pi(c, c') - \Pi(c', c')) - (\Pi(c', c') - \tau(c')) \\
= & \int_c^{c'} \left(\int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(z)) \sigma(\mathrm{d}D|z) + \Pi_1(z, c') \right) \mathrm{d}z \\
= & \int_c^{c'} \left(\int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(z)) \sigma(\mathrm{d}D|z) - \int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(z)) \sigma(\mathrm{d}D|c') \right) \mathrm{d}z \\
= & \int_c^{c'} \left(\int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(z)) (\sigma(\mathrm{d}D|z) - \sigma(\mathrm{d}D|c')) \right) \mathrm{d}z \\
\geq & 0,
\end{aligned}$$

where the second equality follows from assertion 1 and the fundamental theorem of calculus, the last equality follows from (46), and the last inequality follows from assertion 2. As such, the mechanism (σ, τ) is incentive compatible. ■

D.3 Proof of Lemma 2

Proof of Lemma 2. Given any nondecreasing function $\psi \in \mathbb{R}_+^C$, and any ψ-quasi-perfect segmentation $\sigma \in \mathcal{S}$, suppose that for any $c \in C$,

$$\psi(z) \leq \overline{\boldsymbol{p}}_D(z),$$

for Lebesgue almost all $z \in [\underline{c}, c]$ and for all $D \in \mathrm{supp}(\sigma(c))$. Then, for any $c, c' \in C$ with $c < c'$,

$$\begin{aligned}
&\int_c^{c'} \left(\int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(z))(\sigma(\mathrm{d}D|z) - \sigma(\mathrm{d}D|c')) \right) \mathrm{d}z \\
=&\int_c^{c'} \left(\int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(z))\sigma(\mathrm{d}D|z) - \int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(z))\sigma(\mathrm{d}D|c') \right) \mathrm{d}z \\
=&\int_c^{c'} \left(D_0(\psi(z)) - \int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(z))\sigma(\mathrm{d}D|c') \right) \mathrm{d}z \\
\geq&\int_c^{c'} \left(D_0(\psi(z)) - \int_{\mathcal{D}} D(\psi(z))\sigma(\mathrm{d}D|c') \right) \mathrm{d}z \\
=&\int_c^{c'} \left(D_0(\psi(z)) - D_0(\psi(z)) \right) \mathrm{d}z \\
=&0,
\end{aligned}$$

where the second equality follows from assertion 3 of Lemma 11, the inequality follows from (8), and the third equality follows from $\sigma(z) \in \mathcal{S}$ for all $z \in [c, c']$. Together with Lemma 12, this implies that

$$\int_c^{c'} \left(\int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(z))(\sigma(\mathrm{d}D|z) - \sigma(\mathrm{d}D|c')) \right) \mathrm{d}z \geq 0$$

for all $c, c' \in C$. Therefore, by Lemma 1, there exists a transfer τ such that (σ, τ) is incentive compatible, as desired. ∎

D.4 Proof of Lemma 3

Proof of Lemma 3. I first prove the lemma for D_0 being a step function. Consider step function $D \in \mathcal{D}$ and any $\psi \in \mathbb{R}_+^C$ such that $c \leq \psi(c) \leq \overline{\boldsymbol{p}}_D(c)$ for all $c \in C$ and fix any $c \in C$, let

$$V^+ := \{v \in \mathrm{supp}(D) : v \geq \psi(c)\}$$

and let

$$\hat{c} := \inf\{z \in C : \overline{\boldsymbol{p}}_D(z) \geq \psi(c)\}.$$

Since $\overline{\boldsymbol{p}}_D$ is nondecreasing, it then follows $\overline{\boldsymbol{p}}_D(z) \geq \psi(c)$ for all $z \in [\hat{c}, \overline{c}]$ and $\overline{\boldsymbol{p}}_D(z) \leq \psi(c)$ for all $z \in [\underline{c}, \hat{c})$. Moreover, since $\psi(c) \leq \overline{\boldsymbol{p}}_D(c)$, $\hat{c} \leq c$. Furthermore, by definition of $\hat{c}$, it must be either $\hat{c} = \underline{c}$ or $\hat{c} > \underline{c}$ and $\underline{\boldsymbol{p}}_D(\hat{c}) < \psi(c) \leq \overline{\boldsymbol{p}}_D(\hat{c})$, since otherwise, if $\hat{c} > \underline{c}$ and $\underline{\boldsymbol{p}}_D(\hat{c}) \geq \psi(c)$, then for $\varepsilon > 0$ small enough, as $|\mathrm{supp}(D)| < \infty$, $\overline{\boldsymbol{p}}_D(\hat{c} - \varepsilon) = \underline{\boldsymbol{p}}_D(\hat{c}) \geq \psi(c)$, contradicting to the definition of $\hat{c}$. Consider first the case where $\hat{c} > \underline{c}$. In this case, for each $v \in V^+$, define $\hat{m}^v$ recursively as the following

$$\hat{m}^v(v') := \begin{cases} 0, & \text{if } v' \geq \psi(c) \text{ and } v' \neq v \\ m^D(v'), & \text{if } v' = v \\ \beta^*(v|v')m^D(v'), & \text{if } \underline{\boldsymbol{p}}_D(\hat{c}) \leq v' < \psi(c) \\ \alpha^*(v)m^D(v'), & \text{if } v' < \underline{\boldsymbol{p}}_D(\hat{c}) \end{cases}, \forall v' \in \mathrm{supp}(D), \forall v \in V^+,$$

where for all $v \in V^+$ and all $v' \in \mathrm{supp}(D)$ s.t. $\underline{\boldsymbol{p}}_D(\hat{c}) \leq v' < \psi(c)$,

$$\beta^*(v|v') := \frac{(v - \hat{c})m^D(v) - (v' - \hat{c})\sum_{\hat{v} > v'} \hat{m}^v(\hat{v})}{\sum_{v \geq \psi(c)} \left[(v - \hat{c})m^D(v) - (v' - \hat{c})\sum_{\hat{v} > v'} \hat{m}^v(\hat{v})\right]},$$

and for all $v \in V^+$,

$$\alpha^*(v) := \frac{\sum_{\hat{v} \geq \underline{\boldsymbol{p}}_D(\hat{c})} \hat{m}^v(\hat{v})}{\sum_{\hat{v} \geq \underline{\boldsymbol{p}}_D(\hat{c})} m^D(\hat{v})}.$$

By construction,

$$\sum_{v \in V^+} \alpha^*(v) = \sum_{v \in V^+} \beta^*(v|v') = 1 \tag{48}$$

for all $v' \in \mathrm{supp}(D)$ with $\underline{\boldsymbol{p}}_D(\hat{c}) \leq v' < \psi(c)$. As such,

$$\sum_{v \in V^+} \hat{m}^v(v') = m^D(v'), \; \forall v' \in \mathrm{supp}(D). \tag{49}$$

Notice that since $\hat{c} \leq \underline{\boldsymbol{p}}_D(\hat{c}) < \psi(c) \leq \overline{\boldsymbol{p}}_D(\hat{c})$, it must be that

$$\sum_{v \geq \psi(c)} (v - \hat{c}) m^D(v) \geq \sum_{v \geq \overline{\boldsymbol{p}}_D(\hat{c})} (v - \hat{c}) m^D(v) \geq (\overline{\boldsymbol{p}}_D(\hat{c}) - \hat{c}) D(\overline{\boldsymbol{p}}_D(\hat{c})) = (\underline{\boldsymbol{p}}_D(\hat{c}) - \hat{c}) D(\underline{\boldsymbol{p}}_D(\hat{c})). \tag{50}$$

Now consider any $v' \in \mathrm{supp}(D)$ such that $\underline{\boldsymbol{p}}_D(\hat{c}) \leq v' < \psi(c)$. Notice first that

$$\begin{aligned}
&\sum_{v \geq \psi(c)} \left[(v - \hat{c}) m^D(v) - (v' - \hat{c}) \sum_{\hat{v} > v'} \hat{m}^v(\hat{v}) \right] \\
=& \sum_{v \geq \psi(c)} (v - \hat{c}) m^D(v) - (v' - \hat{c}) \sum_{\hat{v} > v'} m^D(\hat{v}) \\
\geq& (\underline{\boldsymbol{p}}_D(\hat{c}) - \hat{c}) D(\underline{\boldsymbol{p}}_D(\hat{c})) - (v' - \hat{c}) \sum_{\hat{v} > v'} m^D(\hat{v}) \\
\geq& (v' - \hat{c}) \sum_{\hat{v} \geq v'} m^D(\hat{v}) - (v' - \hat{c}) \sum_{\hat{v} > v'} m^D(\hat{v}) \\
=& (v' - \hat{c}) m^D(v') \\
\geq& 0,
\end{aligned}$$

where the first equality follows from (49), the first inequality follows from (50), the second inequality follows from the fact that $\underline{\boldsymbol{p}}_D(\hat{c}) \in \boldsymbol{P}_D(\hat{c})$, and the last inequality follows from $\underline{\boldsymbol{p}}_D(\hat{c}) \geq \hat{c}$. As such, for any $v' \in \mathrm{supp}(D)$ with $\underline{\boldsymbol{p}}_D(\hat{c}) \leq v' < \psi(c)$ and for any $v \in V^+$, if

$$(v - \hat{c}) m^D(v) - (v' - \hat{c}) \sum_{\hat{v} > v'} \hat{m}^v(\hat{v}) \geq 0,$$

then $\beta^*(v|v') \geq 0$ and

$$\begin{aligned}
& \hat{m}^v(v') \leq \frac{(v - \hat{c}) m^D(v) - (v' - \hat{c}) \sum_{\hat{v} > v'} \hat{m}^v(\hat{v})}{(v' - \hat{c}) m^D(v')} m^D(v') \\
\iff & (v' - \hat{c}) \hat{m}^v(v') + (v' - \hat{c}) \sum_{\hat{v} > v'} \hat{m}^v(\hat{v}) \leq (v - \hat{c}) m^D(v) \\
\iff & (v' - \hat{c}) \sum_{\hat{v} \geq v'} \hat{m}^v(\hat{v}) \leq (v - \hat{c}) \hat{m}^v(v),
\end{aligned}$$

which in turn implies that

$$(v - \hat{c}) m^D(v) - (v'' - \hat{c}) \sum_{\hat{v} > v''} \hat{m}^v(\hat{v}) > (v - \hat{c}) m^D(v) - (v' - \hat{c}) \sum_{\hat{v} \geq v'} \hat{m}^v(\hat{v}) \geq 0,$$

where $v'' \in \text{supp}(D)$ is the largest element of $\{\hat{v} \in \text{supp}(D) : \underline{\boldsymbol{p}}_D(\hat{c}) \le \hat{v} < v'\}$. Moreover, if $v' = \max\{\hat{v} \in \text{supp}(D) : \underline{\boldsymbol{p}}_D(\hat{c}) \le \hat{v} < \psi(c)\}$, then clearly, for all $v \in V^+$,

$$(v - \hat{c})m^D(v) - \sum_{\hat{v} > v'} \hat{m}^v(v') = (v - v')m^D(v) \ge 0.$$

Therefore, by induction, for any $v' \in \text{supp}(D)$ such that $\underline{\boldsymbol{p}}_D(\hat{c}) \le v' < \psi(c)$, it must be that $\beta^*(v|v') \ge 0$ for all $v \in V^+$ and that

$$(v' - \hat{c}) \sum_{\hat{v} \ge v'} \hat{m}^v(\hat{v}) \le (v - \hat{c})\hat{m}^v(v). \tag{51}$$

Together with (48), this also ensures that

$$\alpha^* \in \Delta(V^+) \tag{52}$$

and

$$\beta^*(v') \in \Delta(V^+), \tag{53}$$

for all $v' \in \text{supp}(D)$ such that $\underline{\boldsymbol{p}}_D(\hat{c}) \le v' < \psi(c)$.

On the other hand, for any $v' \in \text{supp}(D)$ with $v' \le \underline{\boldsymbol{p}}_D(\hat{c})$ and any $v \in V^+$, notice that by the definition of α^*,

$$\sum_{\hat{v} \ge v'} \hat{m}^v(\hat{v}) = \alpha^*(v) \sum_{v' \le \hat{v} < \underline{\boldsymbol{p}}_D(\hat{c})} m^D(\hat{v}) + \sum_{\hat{v} \ge \underline{\boldsymbol{p}}_D(\hat{c})} \hat{m}^v(\hat{v}) = \alpha^*(v) \sum_{\hat{v} \ge v'} m^D(\hat{v}). \tag{54}$$

Thus, for any $v' \in \text{supp}(D)$ with $v' < \underline{\boldsymbol{p}}_D(\hat{c})$ and any $v \in V^+$,

$$\begin{aligned}(v' - \hat{c}) \sum_{\hat{v} \ge v'} \hat{m}^v(\hat{v}) =& \alpha^*(v)(v' - \hat{c})D(v') \\ \le& \alpha^*(v)(\underline{\boldsymbol{p}}_D(\hat{c}) - \hat{c})D(\underline{\boldsymbol{p}}_D(\hat{c})) \\ =& (\underline{\boldsymbol{p}}_D(\hat{c}) - \hat{c}) \sum_{\hat{v} \ge \underline{\boldsymbol{p}}_D(\hat{c})} \hat{m}^v(\hat{v}) \\ \le& (v - \hat{c})\hat{m}^v(v),\end{aligned} \tag{55}$$

where both equalities follow from (54), the first inequality follows from the fact that $\underline{\boldsymbol{p}}_D(\hat{c}) \in \boldsymbol{P}_D(\hat{c})$, and the last inequality follows from (51) by taking $v' = \underline{\boldsymbol{p}}_D(\hat{c})$.

Moreover, by (54), for any $z \in [\underline{c}, \hat{c})$, and any $v \in V^+$, since $\overline{\boldsymbol{p}}_D(z) \le \underline{\boldsymbol{p}}_D(\hat{c})$, it must be that for all $v' \le \overline{\boldsymbol{p}}_D(z)$,

$$\begin{aligned}(v' - z) \sum_{\hat{v} \ge v'} \hat{m}^v(\hat{v}) =& \alpha^*(v)(v' - z)D(v') \\ \le& \alpha^*(v)(\underline{\boldsymbol{p}}_D(z) - z)D(\underline{\boldsymbol{p}}_D(z)) \\ =& (\underline{\boldsymbol{p}}_D(z) - z) \sum_{\hat{v} \ge \underline{\boldsymbol{p}}_D(z)} \hat{m}^v(\hat{v}).\end{aligned} \tag{56}$$

Finally, if $\hat{c} = \underline{c}$, then define $\{\hat{m}^v\}_{v \in V^+}$ as

$$\hat{m}^v(v') := \begin{cases} m^D(v'), & \text{if } v' = v \\ 0, & \text{if } v' \ge \psi(c) \text{ and } v' \ne v \\ \alpha^*(v)m^D(v'), & \text{if } v' < \psi(c) \end{cases}, \forall v' \in V,\, v \in V^+,\, v \ge \overline{\boldsymbol{p}}_D(\underline{c})$$

and

$$\hat{m}^v(v') := \begin{cases} m^D(v'), & \text{if } v' = v \\ 0, & \text{if } v' \neq v \end{cases}, \forall v' \in V,\, v \in V^+,\, \psi(c) \leq v < \overline{\boldsymbol{p}}_D(\underline{c})$$

where

$$\alpha^*(v) := \frac{m^D(v)}{\sum_{v' \geq \overline{\boldsymbol{p}}_D(\underline{c})} m^D(v')}.$$

Again,

$$\sum_{v \geq \overline{\boldsymbol{p}}_D(\underline{c})} \alpha^*(v) = 1 \tag{57}$$

and hence

$$\sum_{v \in V^+} \hat{m}^v(v') = m^D(v'),\, \forall v' \in V. \tag{58}$$

Then, for any $v \geq \overline{\boldsymbol{p}}_D(\underline{c})$ and any $v' \in \text{supp}(D)$ with $v' < \psi(c)$,

$$(v' - \underline{c}) \sum_{\hat{v} \geq v'} \hat{m}^v(v') = \alpha^*(v)(v' - \underline{c})D(v') \leq \alpha^*(v)(\overline{\boldsymbol{p}}_D(\underline{c}) - \underline{c})D(\overline{\boldsymbol{p}}_D(\underline{c})) \leq (v - \underline{c})\hat{m}^v(v). \tag{59}$$

Together, in both of the cases above, from the constructed $\{\hat{m}^v\}_{v \in V^+}$, for each $v \in V^+$, let

$$m^v(v') := \frac{\hat{m}^v(v')}{\sum_{\hat{v} \in V} \hat{m}^v(\hat{v})},\, \forall v' \in \text{supp}(D)$$

and let $D_v(p) := m^v([p, \overline{v}])$. By (52), (53) and (57), in each case, $D_v \in \mathcal{D}$ for all $v \in V^+$. Now define $\sigma(c) \in \Delta^f(\mathcal{D})$ by

$$\sigma(D_v|c) := \sum_{v' \in V} \hat{m}^v(v'),\, \forall v \in V^+.$$

By (49) and (58), in each case, $\sigma(c) \in \mathcal{S}_D$. Furthermore, since m^v is proportional to $\hat{m}^v$ for all $v \in V^+$, (51), (55) and (59) ensure that in each case, $\sigma(c)$ is a $\psi(c)$-quasi-perfect segmentation for $\hat{c}$. Meanwhile, since $\hat{c} \leq c$, $\sigma(c)$ is also a $\psi(c)$-quasi-perfect segmentation for c. Finally, since m^v is proportional to $\hat{m}^v$, (56) implies that for any $z \in [\underline{c}, \hat{c})$,

$$\overline{\boldsymbol{p}}_{D'}(z) \geq \overline{\boldsymbol{p}}_D(z),\, \forall D' \in \text{supp}(\sigma(c)).$$

In addition, by the conclusion that $\sigma(c)$ is a $\psi(c)$-quasi-perfect segmentation for $\hat{c} \leq c$, for any $z \in [\hat{c}, c]$, since $c \leq \psi(c)$ and since $\overline{\boldsymbol{p}}_D$ is nondecreasing for any $D' \in \mathcal{D}$,

$$\overline{\boldsymbol{p}}_{D'}(z) \geq \overline{\boldsymbol{p}}_{D'}(\hat{c}) \geq \psi(c),\, \forall D' \in \text{supp}(\sigma(c)).$$

In particular, $\sigma(c)$ is also a $\psi(c)$-quasi-perfect segmentation for c. Together with the fact that ψ is nondecreasing and that $\psi \leq \overline{\boldsymbol{p}}_D$, it then follows that for any $z \in [\underline{c}, c]$ and for any $D \in \text{supp}(\sigma(c))$, , $\psi(z) \leq \overline{\boldsymbol{p}}_D(z)$. Since $c \in C$ is arbitrary, this ensures that there exists a ψ-quasi-perfect scheme $\sigma \in \mathcal{S}_D^C$ that satisfies (8).

Now consider any $D_0 \in \mathcal{D}$ and any nondecreasing $\psi \in \mathbb{R}_+^C$ with $c \leq \psi(c) \leq \overline{\boldsymbol{p}}_0(c)$ for all $c \in C$. I first construct a sequence of step functions $\{D_n\} \subseteq \mathcal{D}$ such that $\{D_n\} \to D_0$ and that $c \leq \psi(c) \leq \overline{\boldsymbol{p}}_{D_n}(c)$ for all $c \in C$ and for all $n \in \mathbb{N}$. To this end, for each $n \in \mathbb{N}$, first partition V by $\underline{v} = v_0 < v_1 < \ldots < v_n = \overline{v}$ and let $V_k := [v_{k-1}, v_k]$. Then define D_n by $D_n(p) := D_0(v_k)$, for all $p \in V_k$, for all $k \in \{1, \ldots, n\}$ (i.e., by moving all the masses on interval V_k to the top v_k). By construction, it must be that $\overline{\boldsymbol{p}}_{D_n}(c) \geq \overline{\boldsymbol{p}}_0(c)$ for all $c \in C$ and for all $n \in \mathbb{N}$ and hence $c \leq \psi(c) \leq \overline{\boldsymbol{p}}_{D_n}(c)$ for all $c \in C$ and for all $n \in \mathbb{N}$. Also, by construction, $\{D_n\} \to D_0$, as desired.

As such, as shown above, for each $n \in \mathbb{N}$, there exists a ψ-quasi-perfect scheme σ_n such that for all $c \in C$,

$$\psi(z) \leq \overline{\boldsymbol{p}}_D(z)$$

for all $D \in \text{supp}(\sigma_n(c))$ and for all $z \in [\underline{c}, c]$. Furthermore, according to Helly's selection theorem, by possibly taking a subsequence,[25] $\{\sigma_n\} \to \sigma$ for some $\sigma : C \to \Delta(\mathcal{D})$. By Lemma 13, $\sigma \in \mathcal{S}^C$ and is a ψ-quasi-perfect scheme.

It then remains to show that σ satisfies (8). To this end, fix any $c \in C$ and consider any $D \in \text{supp}(\sigma(c))$, by definition, for any $\delta > 0$, $\sigma(N_\delta(D)|c) > 0$.[26] Furthermore, since $\sigma(c)$ has at most countably many atoms, there exists a sequence $\{\delta_k\} \subset (0, 1]$ such that $\{\delta_k\} \to 0$, $\sigma(N_{\delta_k}(D)|c) > 0$ and $\sigma(\partial N_{\delta_k}(D)|c) = 0$ for all $k \in \mathbb{N}$. As a result, since $\{\sigma_n(c)\} \to \sigma(c)$ under the weak-* topology, $\lim_{n\to\infty} \sigma_n(N_{\delta_k}(D)|c) = \sigma(N_{\delta_k}(D)|c) > 0$ for all $k \in \mathbb{N}$. Thus, for each $k \in \mathbb{N}$, there exists $n_k \in \mathbb{N}$ such that $\sigma_{n_k}(N_{\delta_k}(D)|c) > 0$. Moreover, since $\sigma_n(c)$ has finite support as D_n is a step function and $\sigma_n(c) \in \mathcal{S}_{D_n}$, there must be some $D_{n_k} \in N_{\delta_k}(D)$ such that $D_{n_k} \in \text{supp}(\sigma_{n_k}(c))$. Notice that for the subsequence $\{n_k\}$, $\{D_{n_k}\} \to D$ and $D_{n_k} \in \text{supp}(\sigma_{n_k}(c))$ for all $k \in \mathbb{N}$. As a result, together with Lemma 10, since σ_{n_k} satisfies (8) for all $k \in \mathbb{N}$, for Lebesgue almost all $z \in [\underline{c}, c]$,

$$\psi(z) \leq \limsup_{k\to\infty} \overline{\boldsymbol{p}}_{D_{n_k}}(z) \leq \overline{\boldsymbol{p}}_D(z).$$

Since $c \in C$ and $D \in \text{supp}(\sigma(c))$ are arbitrary, this completes the proof. ■

D.5 Proof of Theorem 1

Proof of Theorem 1. I first show that the data broker's optimal revenue must be the same as the price-controlling data broker's optimal revenue R^*. To see this, since $c \leq \overline{\varphi}_G(c) \leq \overline{\boldsymbol{p}}_0(c)$ for all $c \in C$ and $\overline{\varphi}_G \in \mathbb{R}_+^C$ is nondecreasing, by Lemma 3, there exists a $\overline{\varphi}_G$-quasi-perfect scheme $\sigma^* \in \mathcal{S}^C$ that satisfies (8). Together with Lemma 2, there exists a transfer τ^* such that (σ^*, τ^*) is incentive compatible. Moreover, since $\sigma \in \mathcal{S}^C$ is a $\overline{\varphi}_G$-quasi-perfect scheme, by assertion 3 and assertion 4 of Lemma 11, for any $c \in C$,

$$\int_{\mathcal{D}} (\overline{\boldsymbol{p}}_D(c) - \phi_G(c)) D(\overline{\boldsymbol{p}}_D(c)) \sigma^*(\mathrm{d}D|c) = \int_{\{v \geq \overline{\varphi}_G(c)\}} (v - \phi_G(c)) D_0(\mathrm{d}v). \tag{60}$$

[25] See, for instance, Porter (2005) for a generalized version of Helly's selection theorem. To apply this theorem, notice that the family of functions $\{\sigma_n\}$ is of bounded p-variation due to the quasi-perfect structure. Furthermore, for any $c \in C$, the set $\text{cl}(\{\sigma_n(c)\})$ is closed in a compact metric space $\Delta(\mathcal{D})$ and hence is itself compact. As such, there exists a pointwise convergent subsequence of $\{\sigma_n\}$.

[26] $N_\delta(D)$ is the δ-ball around D under the Lévy-Prokhorov metric on $\mathcal{D}$.

Also, by possibly adding a constant to the transfer τ^*, the indirect utility of the producer with cost $\overline{c}$, $U(\overline{c})$, equals to $\bar{\pi}$ under the mechanism (σ^*, τ^*). Thus, for any $c \in C$,

$$\begin{aligned}\int_{\mathcal{D}} \pi_D(c)\sigma^*(\mathrm{d}D|c) - \tau(c) =& U(\overline{c}) + \int_c^{\overline{c}} \left(\int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(z))\sigma^*(\mathrm{d}D|z) \right) \mathrm{d}z \\ =& \bar{\pi} + \int_c^{\overline{c}} D_0(\overline{\varphi}_G(z)) \,\mathrm{d}z \\ \geq& \bar{\pi} + \int_c^{\overline{c}} D_0(\overline{\boldsymbol{p}}_0(z)) \,\mathrm{d}z \\ =& \pi_0(c),\end{aligned}$$

where the first equality follows from Lemma 1, the second equality follows from assertion 3 of Lemma 11, the inequality follows from $\overline{\varphi}_G \leq \overline{\boldsymbol{p}}_0$ and the last equality follows from (5). As a result, (σ^*, τ^*) is individually rational and, together with (60) and Lemma 1,

$$\begin{aligned}\mathbb{E}[\tau^*(c)] =& \int_C \left(\int_{\mathcal{D}} (\overline{\boldsymbol{p}}_D(c) - \phi_G(c)) D(\overline{\boldsymbol{p}}_D(c))\sigma^*(\mathrm{d}D|c) \right) G(\mathrm{d}c) - \bar{\pi} \\ =& \int_C \left(\int_{\{v \geq \overline{\varphi}_G(c)\}} (v - \phi_G(c)) D_0(\mathrm{d}v) \right) G(\mathrm{d}c) - \bar{\pi} \\ =& R^*,\end{aligned}$$

as desired.

Since the data broker's optimal revenue is R^* and since (60) holds for any $\overline{\varphi}_G$-quasi-perfect scheme σ, by Lemma 1, any incentive feasible $\overline{\varphi}_G$-quasi-perfect mechanism must give revenue R^* and hence is optimal.

Conversely, to see why any optimal mechanism must be a $\overline{\varphi}_G$-quasi-perfect mechanism, consider any optimal mechanism (σ, τ). As it is optimal and incentive compatible, by Lemma 1,

$$R^* = \mathbb{E}[\tau(c)] = \int_C \left(\int_{\mathcal{D}} (\boldsymbol{p}_D(c) - \phi_G(c)) D(\boldsymbol{p}_D(c))\sigma(\mathrm{d}D|c) \right) G(\mathrm{d}c) - \bar{\pi}, \tag{61}$$

for any $\boldsymbol{p} \in \boldsymbol{P}$. Also, since (σ, τ) is incentive compatible, for any $\boldsymbol{p} \in \boldsymbol{P}$, the function $\boldsymbol{D}_{\boldsymbol{p}}^{\sigma} \in [0,1]^C$, defined as

$$\boldsymbol{D}_{\boldsymbol{p}}^{\sigma}(c) := \int_{\mathcal{D}} D(\boldsymbol{p}_D(c))\sigma(\mathrm{d}D|c),\ \forall c \in C$$

is nonincreasing.[27] Thus, by (37),

$$\int_C \overline{\phi}_G(c) \left(\int_{\mathcal{D}} D(\boldsymbol{p}_D(c))\sigma(\mathrm{d}D|c) \right) G(\mathrm{d}c) \geq \int_C \overline{\varphi}_G(c) \left(\int_{\mathcal{D}} D(\boldsymbol{p}_D(c))\sigma(\mathrm{d}D|c) \right) G(\mathrm{d}c). \tag{62}$$

Moreover, since (σ, τ) is individually rational, by Lemma 1, it must be that

$$\int_c^{\overline{c}} \left(\int_{\mathcal{D}} D(\boldsymbol{p}_D(z))\sigma(\mathrm{d}D|z) \right) \mathrm{d}z \geq \int_c^{\overline{c}} D_0(\overline{\boldsymbol{p}}_0(z)) \,\mathrm{d}z,\ \forall c \in C. \tag{63}$$

[27]To see this, notice that U is convex since it is a pointwise supremum of convex functions, which is because $\pi_D(c)$ is convex for all D. Lemma 1 implies that the derivative of U is $-\boldsymbol{D}_{\boldsymbol{p}}^{\sigma}$ and thus $\boldsymbol{D}_{\boldsymbol{p}}^{\sigma}$ must be nonincreasing.

Furthermore, since σ^* is a $\overline{\varphi}_G$-quasi-perfect scheme, Lemma 11 implies that, for all $c \in C$,

$$\boldsymbol{D}^{\sigma^*}_{\overline{\boldsymbol{p}}}(c) = \int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(c))\sigma^*(\mathrm{d}D|c) = D_0(\overline{\varphi}_G(c)).$$

Together with (38) and (39), we have

$$\int_C \overline{\varphi}_G(c) D_0(\overline{\varphi}_G(c)) G(\mathrm{d}c) = \int_C \overline{\phi}_G(c) D_0(\overline{\varphi}_G(c)) G(\mathrm{d}c). \tag{64}$$

Now suppose that (σ,τ) is not a $\overline{\varphi}_G$-quasi-perfect mechanism or it does not induce $\overline{\varphi}_G(c)$-quasi-perfect price discrimination for a positive G-measure of c, then there exists $\boldsymbol{p} \in \boldsymbol{P}$, a positive G-measure of c and a positive $\sigma(c)$-measure of $D \in \mathcal{D}$ such that either $\boldsymbol{p}_D(c) < \overline{\boldsymbol{p}}_D(c)$, or $D(c) > 0$ and either $\#\{v \in \mathrm{supp}(D) : v \geq \overline{\varphi}_G(c)\} \neq 1$ or $\max(\mathrm{supp}(D)) \notin \boldsymbol{P}_D(c)$, which imply that there is a positive G-measure of c and a positive $\sigma(c)$-measure of D such that

$$\begin{aligned}
\int_{\{v \geq \overline{\varphi}_G(c)\}} (v - \overline{\varphi}_G(c)) D(\mathrm{d}v) \geq & \int_{\{v \geq \boldsymbol{p}_D(c)\}} (v - \overline{\varphi}_G(c)) D(\mathrm{d}v) \\
=& (\boldsymbol{p}_D(c) - \overline{\varphi}_G(c)) D(\boldsymbol{p}_D(c)) + \int_{\{v \geq \boldsymbol{p}_D(c)\}} (v - \boldsymbol{p}_D(c)) D(\mathrm{d}v) \\
\geq & (\boldsymbol{p}_D(c) - \overline{\varphi}_G(c)) D(\boldsymbol{p}_D(c)),
\end{aligned}$$

with at least one inequality being strict. Therefore,

$$\int_C \left(\int_{\mathcal{D}} (\boldsymbol{p}_D(c) - \overline{\varphi}_G(c)) D(\boldsymbol{p}_D(c)) \sigma(\mathrm{d}D|c) \right) G(\mathrm{d}c) < \int_C \left(\int_V (v - \overline{\varphi}_G(c))^+ D_0(\mathrm{d}v) \right) G(\mathrm{d}c). \tag{65}$$

Moreover, by (61), since

$$\begin{aligned}
&\int_C \left(\int_{\mathcal{D}} (\boldsymbol{p}_D(c) - \overline{\varphi}_G(c)) D(\boldsymbol{p}_D(c)) \sigma(\mathrm{d}D|c) \right) G(\mathrm{d}c) \\
&\quad + \int_C (\overline{\varphi}_G(c) - \phi_G(c)) \left(\int_{\mathcal{D}} D(\boldsymbol{p}_D(c)) \sigma(\mathrm{d}D|c) \right) G(\mathrm{d}c) \\
=& \int_{\mathcal{D}} \left(\int_{\mathcal{D}} (\boldsymbol{p}_D(c) - \phi_G(c)) D(\boldsymbol{p}_D(c)) \sigma(\mathrm{d}D|c) \right) G(\mathrm{d}c) \\
=& \int_C \left(\int_{\{v \geq \overline{\varphi}_G(c)\}} (v - \phi_G(c)) D_0(\mathrm{d}v) \right) G(\mathrm{d}c) \\
=& \int_C \left(\int_V (v - \overline{\varphi}_G(c))^+ D_0(\mathrm{d}v) \right) G(\mathrm{d}c) + \int_C (\overline{\varphi}_G(c) - \phi_G(c)) D_0(\overline{\varphi}_G(c)) G(\mathrm{d}c),
\end{aligned}$$

together with, (65), it must be that

$$\begin{aligned}
&\int_C (\overline{\phi}_G(c) - \phi_G(c)) \left(\int_{\mathcal{D}} D(\boldsymbol{p}_D(c)) \sigma(\mathrm{d}D|c) \right) G(\mathrm{d}c) \\
\geq & \int_C (\overline{\varphi}_G(c) - \phi_G(c)) \left(\int_{\mathcal{D}} D(\boldsymbol{p}_D(c)) \sigma(\mathrm{d}D|c) \right) G(\mathrm{d}c) \\
> & \int_C (\overline{\varphi}_G(c) - \phi_G(c)) D_0(\overline{\varphi}_G(c)) G(\mathrm{d}c) \\
= & \int_C (\overline{\phi}_G(c) - \phi_G(c)) D_0(\overline{\varphi}_G(c)) G(\mathrm{d}c),
\end{aligned}$$

where the first inequality follows from (62) and the equality follows from (64). Furthermore, since $\overline{\phi}_G(c) = \phi_G(c)$ for all $c \in [\underline{c}, c^*]$ and $\overline{\phi}_G(c) = \overline{\varphi}_G(c) = \overline{\boldsymbol{p}}_0(c)$ for all $c \in (c^*, \overline{c}]$, it then follows that

$$\int_{c^*}^{\overline{c}} (\phi_G(c) - \overline{\boldsymbol{p}}_0(c)) \left(\int_{\mathcal{D}} D(\boldsymbol{p}_D(c))\sigma(\mathrm{d}D|c) \right) G(\mathrm{d}c) < \int_{c^*}^{\overline{c}} (\phi_G(c) - \overline{\boldsymbol{p}}_0(c)) D_0(\overline{\boldsymbol{p}}_0(c)) G(\mathrm{d}c),$$

Using integration by parts, this is equivalent to

$$\int_{c^*}^{\overline{c}} \left(\int_{c}^{\overline{c}} \left(\int_{\mathcal{D}} D(\boldsymbol{p}_D(z))\sigma(\mathrm{d}D|z) \right) \mathrm{d}z \right) M^*(\mathrm{d}c) < \int_{c^*}^{\overline{c}} \left(\int_{c}^{\overline{c}} D_0(\overline{\boldsymbol{p}}_0(z))\, \mathrm{d}z \right) M^*(\mathrm{d}c),$$

where M^* is defined in (34). However, by (63) and by the fact that M^* is a CDF of a Borel measure, which is due to Assumption 1,

$$\int_{c^*}^{\overline{c}} \left(\int_{c}^{\overline{c}} \left(\int_{\mathcal{D}} D(\boldsymbol{p}_D(z))\sigma(\mathrm{d}D|z) \right) \mathrm{d}z \right) M^*(\mathrm{d}c) \geq \int_{c^*}^{\overline{c}} \left(\int_{c}^{\overline{c}} D_0(\overline{\boldsymbol{p}}_0(z))\, \mathrm{d}z \right) M^*(\mathrm{d}c),$$

a contradiction. Therefore, σ must be a $\overline{\varphi}_G$-quasi-perfect scheme and must induce $\overline{\varphi}_G(c)$-quasi-perfect price discrimination for G-almost all $c \in C$. Together with Lemma 1, and the fact that $U(\overline{c}) = \bar{\pi}$ under any optimal mechanism, (σ, τ) must be a $\overline{\varphi}_G$-quasi-perfect mechanism. This completes the proof. ■

D.6 Proof of Theorem 2

Proof of Theorem 2. Notice that from the proof of Lemma 3, for any step function $D \in \mathcal{D}$ that is regular, by replacing ψ with $\overline{\varphi}_G$, the resulting scheme $\sigma \in \mathcal{S}_D^C$ must take form of

$$\sigma\left(D_v^{\overline{\varphi}_G(c)} \middle| c \right) = \frac{m^D(v)}{D(\overline{\varphi}_G(c))} \tag{66}$$

for all $c \in C$ and for all $v \in [\overline{\varphi}_G(c), \overline{v}]$, where $D_v^{\overline{\varphi}_G(c)}$ is defined by (2) with D_0 being replaced by D.

Now, for any regular $D_0 \in \mathcal{D}$, take a sequence of regular step functions $\{D_n\} \subseteq \mathcal{D}$ such that $\{D_n\} \to D_0$ and take the associated segmentation scheme σ_n defined by (66) for each $n \in \mathbb{N}$. By the proof of Lemma 3, $\{\sigma_n\} \to \sigma$ for some $\overline{\varphi}_G$-quasi-perfect scheme $\sigma \in \mathcal{S}^C$ that satisfies (8). Moreover, by the same argument as in the proof of Lemma 3, for all $c \in C$ and for all $D \in \mathrm{supp}(\sigma(c))$, there exists a subsequence $\{D_{n_k}\}$ such that $D_{n_k} \in \mathrm{supp}(\sigma_{n_k}(c))$ and $\{D_{n_k}\} \to D$. This then implies that $D(p) = D_0(p)$ for every $p \in [\underline{v}, \overline{\varphi}_G(c))$ at which D_0 is continuous. Finally, since $\sigma(c)$ is a $\overline{\varphi}_G(c)$-quasi-perfect segmentation for c, it must be that for any $p \in [\overline{\varphi}_G(c), \overline{v}]$, $D(p) \in \{D_0(\overline{\varphi}_G(c)), 0\}$ and for any $p \in [\underline{v}, \overline{\varphi}_G(c))$, $D(p) = D_0(p)$, which means that $D = D_v^{\overline{\varphi}_G(c)}$ for some $v \in [\overline{\varphi}_C(c), \overline{v}]$ and

$$\sigma\left(\left\{ D_v^{\overline{\varphi}_G(c)} : v \geq p \right\} \middle| c \right) = \frac{D_0(p)}{D_0(\overline{\varphi}_G(c))} \tag{67}$$

for any $p \in [\overline{\varphi}_G(c), \overline{v}]$, where $D_v^{\overline{\varphi}_G(c)}$ is defined by (2)

Finally, by the proof of Lemma 3, σ is a $\overline{\varphi}_G$-quasi-perfect scheme satisfying (8) and hence, by Lemma 2, there exists a transfer scheme τ such that (σ, τ) is incentive feasible. Thus, by Theorem 1, (σ, τ) is optimal, and by (67), (σ, τ) is the canonical $\overline{\varphi}_G$-quasi perfect mechanism. ■

E Proofs of Other Main Results

E.1 Proof of Theorem 3

Proof of Theorem 3. Let (σ, τ) be any optimal mechanism. By Theorem 1, (σ, τ) must be a $\overline{\varphi}_G$-quasi-perfect mechanism and induces $\overline{\varphi}_G$-quasi-perfect price discrimination. Therefore, for G-almost all $c \in C$ and for $\sigma(c)$-almost all $D \in \mathcal{D}$, $D(p) = 0$ for any $p > \overline{\boldsymbol{p}}_D(c) = \max(\mathrm{supp}(D))$ and consumer surplus under (σ, τ) is

$$\int_{\mathcal{D}} \left(\int_{\{v \geq \overline{\boldsymbol{p}}_D(c)\}} (v - \overline{\boldsymbol{p}}_D(c)) D(\mathrm{d}v) \right) \sigma^*(\mathrm{d}D|c).$$

Using integration by parts, it then follows that

$$\int_{\mathcal{D}} \left(\int_{\{v \geq \overline{\boldsymbol{p}}_D(c)\}} (v - \overline{\boldsymbol{p}}_D(c)) D(\mathrm{d}v) \right) \sigma^*(\mathrm{d}D|c) = \int_{\mathcal{D}} \left(\int_{\overline{\boldsymbol{p}}_D(c)}^{\overline{v}} D(z) \, \mathrm{d}z \right) \sigma^*(\mathrm{d}D|c) = 0.$$

for G-almost all $c \in C$. Together,

$$\begin{aligned}
&\int_C \left(\int_{\mathcal{D}} \left(\int_{\{v \geq \overline{\boldsymbol{p}}_D(c)\}} (v - \overline{\boldsymbol{p}}_D(c)) D(\mathrm{d}v) \right) \sigma^*(\mathrm{d}D|c) \right) G(\mathrm{d}c) \\
=&\int_C \left(\int_{\mathcal{D}} \left(\int_{\overline{\boldsymbol{p}}_D(c)}^{\overline{v}} D(z) \, \mathrm{d}z \right) \sigma^*(\mathrm{d}D|c) \right) G(\mathrm{d}c) \\
=&0.
\end{aligned}$$

This completes the proof. ∎

E.2 Proof of Proposition 1

Proof of Proposition 1. By Theorem 1, the data broker's optimal revenue is

$$R^* = \int_C \left(\int_{\{v \geq \overline{\varphi}_G(c)\}} (v - \phi_G(c)) D_0(\mathrm{d}v) \right) G(\mathrm{d}c) - \bar{\pi}.$$

On the other hand, since $\boldsymbol{P}_0(c)$ is a singleton for (Lebesgue)-almost all $c \in C$ and since G is absolutely continuous, consumer surplus under uniform pricing is

$$\int_C \left(\int_{\{v \geq \overline{\boldsymbol{p}}_0(c)\}} (v - \overline{\boldsymbol{p}}_0(c)) D_0(\mathrm{d}v) \right) G(\mathrm{d}c).$$

Furthermore, for any $c \in C$,

$$\int_{\{v \geq \overline{\varphi}_G(c)\}} (v - \phi_G(c)) D_0(\mathrm{d}v) = \int_{\{v \geq \overline{\boldsymbol{p}}_0(c)\}} (v - \phi_G(c)) D_0(\mathrm{d}v) + \int_{[\overline{\varphi}_G(c), \overline{\boldsymbol{p}}_0(c))} (v - \phi_G(c)) D_0(\mathrm{d}v). \tag{68}$$

Notice first that for any $c \in C$,

$$
\begin{aligned}
&\int_{\{v \geq \overline{\boldsymbol{p}}_0(c)\}} (v - \phi_G(c)) D_0(\mathrm{d}v) - \int_{\{v \geq \overline{\boldsymbol{p}}_0(c)\}} (v - \overline{\boldsymbol{p}}_0(c)) D_0(\mathrm{d}v) - \bar{\pi} \\
=&(\overline{\boldsymbol{p}}_0(c) - \phi_G(c)) D_0(\overline{\boldsymbol{p}}_0(c)) - \bar{\pi} \\
=&\pi_0(c) - \frac{G(c)}{g(c)} D_0(\overline{\boldsymbol{p}}_0(c)) - \bar{\pi} \\
=&\int_c^{\overline{c}} D_0(\overline{\boldsymbol{p}}_0(z))\, \mathrm{d}z - \frac{G(c)}{g(c)} D_0(\overline{\boldsymbol{p}}_0(c)),
\end{aligned}
$$

and thus

$$
\begin{aligned}
&\int_C \left(\int_{\{v \geq \overline{\boldsymbol{p}}_0(c)\}} (v - \phi_G(c)) D_0(\mathrm{d}v) - \int_{\{v \geq \overline{\boldsymbol{p}}_0(c)\}} (v - \overline{\boldsymbol{p}}_0(c)) D_0(\mathrm{d}v) \right) G(\mathrm{d}c) - \bar{\pi} \\
=&\int_C \left(\int_c^{\overline{c}} D_0(\overline{\boldsymbol{p}}_0(z))\, \mathrm{d}z - \frac{G(c)}{g(c)} D_0(\overline{\boldsymbol{p}}_0(c)) \right) G(\mathrm{d}c) \qquad (69) \\
=&\int_C G(c)(D_0(\overline{\boldsymbol{p}}_0(c)) - D_0(\overline{\boldsymbol{p}}_0(c)))\, \mathrm{d}c \\
=&0.
\end{aligned}
$$

On the other hand, for any $c \in C$,

$$
\begin{aligned}
&\int_{[\overline{\varphi}_G(c), \overline{\boldsymbol{p}}_0(c))} (v - \phi_G(c)) D_0(\mathrm{d}v) \\
=&\int_{[\overline{\varphi}_G(c), \overline{\boldsymbol{p}}_0(c))} (v - \varphi_G(c)) D_0(\mathrm{d}v) + (\varphi_G(c) - \phi_G(c))(D_0(\overline{\varphi}_G(c)) - D_0(\overline{\boldsymbol{p}}_0(c))),
\end{aligned}
$$

and thus,

$$
\begin{aligned}
&\int_C \left(\int_{[\overline{\varphi}_G(c), \overline{\boldsymbol{p}}_0(c))} (v - \phi_G(c)) D_0(\mathrm{d}v) \right) G(\mathrm{d}c) \\
=&\int_C \left(\int_{[\overline{\varphi}_G(c), \overline{\boldsymbol{p}}_0(c))} (v - \varphi_G(c)) D_0(\mathrm{d}v) \right) G(\mathrm{d}c) \\
&+ \int_C (\varphi_G(c) - \phi_G(c))(D_0(\overline{\varphi}_G(c)) - D_0(\overline{\boldsymbol{p}}_0(c))) G(\mathrm{d}c) \\
\geq&\int_C (\varphi_G(c) - \phi_G(c))(D_0(\overline{\varphi}_G(c)) - D_0(\overline{\boldsymbol{p}}_0(c))) G(\mathrm{d}c) \qquad (70) \\
=&\int_C (\varphi_G(c) - \phi_G(c)) D_0(\overline{\varphi}_G(c)) G(\mathrm{d}c) - \int_C (\varphi_G(c) - \phi_G(c)) D_0(\overline{\boldsymbol{p}}_0(c))) G(\mathrm{d}c) \\
\geq&0,
\end{aligned}
$$

where the first inequality follows from $\overline{\varphi}_G = \min\{\varphi_G, \overline{\boldsymbol{p}}_0\}$ and the second inequality follows from the fact that $\overline{\boldsymbol{p}}_0$ is nondecreasing, (37) and (38). Together, combining (68), (69) and (70),

$$\begin{aligned}
&\int_C \left(\int_{\{v \geq \overline{\varphi}_G(c)\}} (v - \phi_G(c)) D_0(\mathrm{d}v)\right) G(\mathrm{d}c) - \bar{\pi} - \int_C \left(\int_{\{v \geq \overline{\boldsymbol{p}}_0(c)\}} (v - \overline{\boldsymbol{p}}_0(c)) D_0(\mathrm{d}v)\right) G(\mathrm{d}c) \\
=& \int_C \left(\int_{\{v \geq \overline{\boldsymbol{p}}_0(c)\}} (v - \phi_G(c)) D_0(\mathrm{d}v) - \int_{\{v \geq \overline{\boldsymbol{p}}_0(c)\}} (v - \overline{\boldsymbol{p}}_0(c)) D_0(\mathrm{d}v)\right) G(\mathrm{d}c) - \bar{\pi} \\
&+ \int_C \left(\int_{[\overline{\varphi}_G(c), \overline{\boldsymbol{p}}_0(c))} (v - \phi_G(c)) D_0(\mathrm{d}v)\right) G(\mathrm{d}c) \\
\geq& 0,
\end{aligned}$$

as desired. ■

E.3 Proof of Lemma 4

To prove Lemma 4, first notice that by the revelation principle, it is without loss to restrict attention to the collection of incentive feasible mechanisms $(\boldsymbol{q}, t)$, where $\boldsymbol{q}(c)$ stands for the quantity purchased for each report $c \in C$ and $t(c)$ stands for the amount of payment from the exclusive retailer to the producer for each report $c \in C$. $(\boldsymbol{q}, t)$ is incentive compatible if for any $c, c' \in C$,

$$t(c) - c\boldsymbol{q}(c) \geq t(c') - c\boldsymbol{q}(c') \tag{IC**}$$

and is individually rational if for any $c \in C$,

$$t(c) - c\boldsymbol{q}(c) \geq \pi_{D_0}(c).$$

The exclusive retailer's problem is then to choose $(\boldsymbol{q}, t)$ to maximize

$$\int_C [\bar{R}(\boldsymbol{q}(c)) - t(c)] G(\mathrm{d}c)$$

subject to (IC**) and (IR**), where $\bar{R}$ is defined by (26).

Proof of Lemma 4. Consider the exclusive retailer's problem. First notice that by standard arguments, $(\boldsymbol{q}, t)$ is incentive compatible if and only if $\boldsymbol{q}$ is nonincreasing and there exists a constant $\bar{t}$ such that

$$t(c) = c\boldsymbol{q}(c) + \int_c^{\overline{c}} \boldsymbol{q}(z)\,\mathrm{d}z - \bar{t},$$

for all $c \in C$. Moreover, any incentive compatible mechanism must give the producer indirect utility

$$U(\overline{c}) + \int_c^{\overline{c}} \boldsymbol{q}(z)\,\mathrm{d}z$$

when her cost is $c \in C$. Together, the exclusive retailer's profit maximization problem can be wirtten as

$$\begin{aligned}
&\max_{\boldsymbol{q} \in \mathcal{Q}} \int_C [\bar{R}(\boldsymbol{q}(c)) - \phi_G(c)\boldsymbol{q}(c)] G(\mathrm{d}c) \\
&\text{s.t. } \bar{\pi} + \int_c^{\overline{c}} \boldsymbol{q}(z)\,\mathrm{d}z \geq \bar{\pi} + \int_c^{\overline{c}} D_0(\overline{\boldsymbol{p}}_0(z))\,\mathrm{d}z.
\end{aligned} \tag{71}$$

Furthermore, by Lemma 15, the objective function of (71) can be written as

$$\int_0^{\boldsymbol{q}(c)} (D_0^{-1}(q) - \phi_G(c)) G(\mathrm{d}c).$$

Thus, together with Lemma 16, the exclusive retailer's profit maximization problem is equivalent to the price-controlling data broker's revenue maximization problem. This completes the proof. ■

E.4 Proof of Theorem 7

Proof of Theorem 7. By Lemma 4, it suffices to prove outcome-equivalence between data brokership and price-controlling data brokership. By Proposition 2 and Theorem 1, both the exclusive retailer and the data broker have optimal revenue R^*. Furthermore, for any optimal mechanism (σ, τ) of the data broker and any optimal mechanism $(\hat{\sigma}, \hat{\tau}, \hat{\boldsymbol{\gamma}})$ of the price-controlling data broker, both of them must induce $\overline{\varphi}_G(c)$-quasi-perfect price discrimination for G-almost all $c \in C$. In particular, for G-almost all $c \in C$, all the consumers with $v \geq \overline{\varphi}_G(c)$ buys the product by paying their values and all the consumers with $v < \overline{\varphi}_G(c)$ do not buy the product. Thus, the consumer surplus and the allocation of the product induced by (σ, τ) and $(\hat{\sigma}, \hat{\tau}, \hat{\boldsymbol{\gamma}})$ are the same.

In addition, for the data broker's optimal mechanism (σ, τ), Theorem 1 implies that σ must be a $\overline{\varphi}_G$-quasi-perfect scheme and hence by Lemma 11, for G-almost all $c \in C$,

$$\int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(c)) \sigma(\mathrm{d}D|c) = D_0(\overline{\varphi}_G(c))$$

and

$$\int_{\mathcal{D}} (\overline{\boldsymbol{p}}_D(c) - \phi_G(c)) D(\overline{\boldsymbol{p}}_D(c)) \sigma(\mathrm{d}D|c) = \int_{\{v \geq \overline{\varphi}_G(c)\}} (v - \phi_G(c)) D_0(\mathrm{d}v).$$

Therefore, for Lebesgue almost all $c \in C$, by Lemma 1,

$$\begin{aligned}\int_{\mathcal{D}} \pi_D(c) \sigma(\mathrm{d}D|c) - \tau(c) =& \overline{\pi} + \int_c^{\overline{c}} \left(\int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(z)) \sigma(\mathrm{d}D|z) \right) \mathrm{d}z \\ =& \overline{\pi} + \int_c^{\overline{c}} D_0(\overline{\varphi}_G(z)) \, \mathrm{d}z. \end{aligned} \tag{72}$$

On the other hand, for the price-controlling data broker's optimal mechanism $(\hat{\sigma}, \hat{\tau}, \hat{\boldsymbol{\gamma}})$, since it induces $\overline{\varphi}_G(c)$-quasi-perfect price discrimination for almost all $c \in C$, it must be that,

$$\int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} D(p) \hat{\boldsymbol{\gamma}}(\mathrm{d}p|D, c) \right) \hat{\sigma}(\mathrm{d}D|c) = D_0(\overline{\varphi}_G(c)),$$

for G-almost all $c \in C$. Together with (44), we have

$$\int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} (p - \phi_G(c)) D(p) \hat{\boldsymbol{\gamma}}(\mathrm{d}p|D, c) \right) \hat{\sigma}(\mathrm{d}D|c) = \int_{\{v \geq \overline{\varphi}_G(c)\}} (v - \phi_G(c)) D_0(\mathrm{d}v).$$

Therefore, by Lemma 14, for any $c \in C$,

$$\begin{aligned} & \int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} (p - c) D(p) \hat{\boldsymbol{\gamma}}(\mathrm{d}p|D, c) \right) \hat{\sigma}(\mathrm{d}D|c) - \hat{\tau}(c) \\ =& \overline{\pi} + \int_c^{\overline{c}} \left(\int_{\mathcal{D}} \left(\int_{\mathbb{R}_+} D(p) \hat{\boldsymbol{\gamma}}(\mathrm{d}p|D, z) \right) \hat{\sigma}(\mathrm{d}D|z) \right) \mathrm{d}z \\ =& \overline{\pi} + \int_c^{\overline{c}} D_0(\overline{\varphi}_G(z)) \, \mathrm{d}z. \end{aligned}$$

Thus, the producer's profit under both (σ,τ) and $(\hat{\sigma},\hat{\tau},\hat{\gamma})$ are the same. This completes the proof. ■

F Proofs for Extensions

F.1 Proof of Theorem 8

Proof of Theorem 8. For each $\theta \in \Theta$, write $\text{supp}(D_\theta)$ as $[l(\theta), u(\theta)]$. Also, for any $p \in V$, let θ_p be the unique θ such that $p \in (l(\theta), u(\theta)]$. Notice that since $\{(l(\theta), u(\theta))\}_{\theta\in\Theta}$ is disjoint, for any $\beta \in \Delta(\Theta)$, any $\theta \in \Theta$, and any $p \in \text{supp}(D_\beta)$,

$$D_\beta(p) = \sum_{\{\theta':u(\theta')\geq u(\theta_p)\}} D_{\theta'}(p)\beta(\theta') = D_\theta(p)\beta(\theta) + \sum_{\{\theta':u(\theta')>u(\theta_p)\}} \beta(\theta').$$

In particular, different prices set in $\text{supp}(D_\theta)$ does not affect the probability of trade through $\theta' \in \Theta$ such that $u(\theta') > u(\theta)$.

As a result, the construction in the proof of Lemma 3 is still valid, with the demands being replaced by D_β. Specifically, for any $\beta \in \Delta(\Theta)$ and any $c \in C$, there exists $\{\beta_i\}_{i=1}^n \subseteq \Delta^f(V)$ such that:

1. $\beta \in \text{co}(\{\beta_i\}_{i=1}^n)$.

2. For each $i \in \{1,\dots,n\}$, the set

$$\{\theta \in \text{supp}(\beta_i) | u(\theta) \geq \overline{\boldsymbol{p}}_{D_{\beta_i}}(c)\}$$

   is nonempty and is a singleton.

3. For each $i \in \{1,\dots,n\}$,

$$\boldsymbol{P}_{D_{\beta_i}}(c) \bigcap \text{supp}(D_{\bar{\theta}_{\beta_i}}) \neq \emptyset,$$

   where $\bar{\theta}_{\beta_i} := \max\{u(\theta) : \theta \in \text{supp}(\beta_i)\}$.

4. For each $i \in \{1,\dots,n\}$ and any $z \in [\underline{c}, c]$,

$$\overline{\boldsymbol{p}}_{D_{\beta_i}}(z) \geq \overline{\boldsymbol{p}}_{D_\beta}(z).$$

This further implies that, by Lemma 11, and by the same argument as in the proof of Lemma 2, for any $\beta \in \Delta(\Theta)$, there exists $\sigma^\beta \in \Delta^f(\Delta(\Theta))^C$ such that

5. For any $c \in C$,

$$\begin{aligned}&\sum_{\beta'\in\text{supp}(\sigma^\beta(c))} (\overline{\boldsymbol{p}}_{D_{\beta'}}(c) - \overline{\boldsymbol{p}}_{D_\beta}(c))D_{\beta'}(\overline{\boldsymbol{p}}_{D'_\beta}(c))\sigma^\beta(\beta'|c)\\ =& \sum_{\{\theta:u(\theta)\geq\theta(\overline{\boldsymbol{p}}_{D_\beta}(c))\}} (\overline{\boldsymbol{p}}_{D_\theta}(c) - \overline{\boldsymbol{p}}_{D_\beta}(c))D_\theta(\overline{\boldsymbol{p}}_{D_\theta}(c))\beta(\theta).\end{aligned}$$

6. For any $c \in C$, $\sum_{\beta'\in\text{supp}(\sigma^\beta(c))} D_{\beta'}(\overline{\boldsymbol{p}}_{D_{\beta'}}(c))\sigma^\beta(\beta'|c) = D_\beta(\overline{\boldsymbol{p}}_{D_\beta}(c))$.

7. For any $c \in C$, $\sum_{\beta'\in\text{supp}(\sigma^\beta(c))} \beta'\sigma^\beta(\beta'|c) = \beta$.

8. For any $\beta' \in \text{supp}(\sigma^\beta(c'))$,
$$\overline{\boldsymbol{p}}_{D_{\beta'}}(c) \geq \overline{\boldsymbol{p}}_{D_\beta}(c),$$
for any $c, c' \in C$ such that $c < c'$ and
$$\sum_{\beta' \in \text{supp}(\sigma^\beta(c))} D_{\beta'}(\overline{\boldsymbol{p}}_{D_{\beta'}}(c))\sigma^\beta(\beta'|c) \geq D(\overline{\boldsymbol{p}}_{D_\beta}(c)),$$
for any $c, c' \in C$ such that $c > c'$.

Now consider any mechanism (σ, τ). Suppose that there is a selection $\boldsymbol{p} \in \boldsymbol{P}$ and a positive G-measure of c such that there exists some $\beta \in \text{supp}(\sigma(c))$ and with
$$\{\theta \in \text{supp}(\beta) : u(\theta) > u(\theta_{\boldsymbol{p}_{D_\beta}(c)})\} \neq \emptyset. \tag{73}$$

Then, for such $\boldsymbol{p} \in \boldsymbol{P}$, $c \in C$ and $\beta \in \text{supp}(\sigma(c))$, there exists $\sigma^\beta(c) \in \Delta^f(\Delta(\Theta))$ such that assertions 5 through 8 above hold. In particular, assertions 5 and 6 imply that
$$\begin{aligned}
&\sum_{\beta' \in \text{supp}(\sigma^\beta(c))} (\overline{\boldsymbol{p}}_{D_{\beta'}}(c) - \phi_G(c))D_{\beta'}(\overline{\boldsymbol{p}}_{D_{\beta'}}(c))\sigma^\beta(\beta'|c)\\
=&\sum_{\beta' \in \text{supp}(\sigma^\beta(c))} (\overline{\boldsymbol{p}}_{D_{\beta'}}(c) - \overline{\boldsymbol{p}}_{D_\beta}(c))D_{\beta'}(\overline{\boldsymbol{p}}_{D_{\beta'}}(c))\sigma^\beta(\beta'|c) + (\overline{\boldsymbol{p}}_{D_\beta}(c) - \phi_G(c))D_\beta(\overline{\boldsymbol{p}}_{D_\beta}(c))\\
\geq&\sum_{\{\theta : u(\theta) \geq u(\theta_{\overline{\boldsymbol{p}}_{D_\beta}(c)})\}} (\overline{\boldsymbol{p}}_{D_\theta}(c) - \overline{\boldsymbol{p}}_{D_\beta}(c))D_\theta(\overline{\boldsymbol{p}}_{D_\theta}(c))\beta(\theta) + (\boldsymbol{p}_{D_\beta}(c) - \phi_G(c))D_\beta(\boldsymbol{p}_{D_\beta}(c))\\
>&(\boldsymbol{p}_{D_\beta}(c) - \phi_G(c))D_\beta(\boldsymbol{p}_{D_\beta}(c)),
\end{aligned}$$
where the second equality follows from 5 and 6 and the inequality is strict due to (73).

As such, together with assertion 7, $\sigma^\beta(c)$ induces another segmentation $\hat{\sigma}(c)$ through
$$\hat{\sigma}(\hat{\beta}|c) := \sum_{\beta \in \text{supp}(\sigma(c))} \sigma^\beta(\hat{\beta}|c)\sigma(\beta|c), \ \forall \hat{\beta} \in \bigcup_{\beta \in \text{supp}(\sigma(c))} \text{supp}(\sigma^\beta(c))$$

As (73) holds with positive G-measure of $c \in C$, the induced segmentation scheme $\hat{\sigma} \in \Delta^f(\Delta(\Theta))^C$ strictly improves the data broker's revenue. Finally, as argued in the proof of Theorem S1 in the Supplemental Material, assertions 6 and 8 above and Lemma 1 ensure that there exists a transfer $\hat{\tau}$ such that $(\hat{\sigma}, \hat{\tau})$ is incentive compatible and individually rational and strictly improves the data broker's revenue.

Together, any optimal mechanism (σ, τ) must be such that for G-almost all $c \in C$ and for all $\beta \in \text{supp}(\sigma(c))$,
$$\{\theta \in \text{supp}(\beta) : u(\theta) > u(\theta_{\boldsymbol{p}_{D_\beta}(c)})\} = \emptyset.$$
which, together with the fact that $\sum_{\beta \in \text{supp}(\sigma(c))} \sigma(\beta|c) = \beta_0$ for all $c \in C$, implies that for G-almost all $c \in C$, the consumer surplus must be lower than the case when all the information about θ is revealed. ■

F.2 Proof of Theorem 9

To prove Theorem 9, I first introduce two useful lemmas.

Lemma 17. *For any $c \in C$, any $\nu \geq c$ and any segmentation $s \in \Delta^f(\Delta(\Theta))$,*

$$\int_{\Delta(\Theta)} (\overline{\boldsymbol{p}}_{D_\beta}(c) - \nu) D_\beta(\overline{\boldsymbol{p}}_{D_\beta}(c)) s(\mathrm{d}\beta) \leq \int_{\{\theta : \overline{\boldsymbol{p}}_{D_\theta}(c) \geq \nu\}} (\overline{\boldsymbol{p}}_{D_\theta}(c) - \nu) D_\theta(\overline{\boldsymbol{p}}_{D_\theta}) \beta_0(\mathrm{d}\theta),$$

Proof. I first show that for any segmentation $s \in \Delta^f(\Delta(\Theta))$, there must exist another segmentation $\hat{s}$ such that for any $\beta \in \mathrm{supp}(\hat{s})$, either $\beta(\{\theta : u(\theta) < c\}) = 1$ or $\overline{\boldsymbol{p}}_{D_\beta}(c) = \overline{\boldsymbol{p}}_{D_{\bar{\theta}_\beta}}(c)$ and

$$\int_{\Delta(\Theta)} (\overline{\boldsymbol{p}}_{D_\beta} - \nu) D(\overline{\boldsymbol{p}}_\beta(c)) s(\mathrm{d}\beta) \leq \int_{\Delta(\Theta)} (\overline{\boldsymbol{p}}_{D_\beta} - \nu) D(\overline{\boldsymbol{p}}_\beta(c)) \hat{s}(\mathrm{d}\beta),$$

where $\bar{\theta}_\beta := \max(\mathrm{supp}(\beta))$. Indeed, consider any segmentation $s \in \Delta^f(\Delta(\Theta))$. For any $\beta \in \mathrm{supp}(s)$, by definition, it must be that $\mathrm{supp}(\beta) \cap \{\theta \in \Theta : u(\theta) \geq \overline{\boldsymbol{p}}_{D_\beta}(c)\} \neq \emptyset$. Now define $\hat{\beta}^\theta$ as

$$\hat{\beta}^\theta(\theta') := \begin{cases} \beta(\theta), & \text{if } \theta' \leq \overline{\boldsymbol{p}}_{D_\beta}(c) \\ \sum_{\{\hat{\theta} : u(\hat{\theta}) \geq \overline{\boldsymbol{p}}_{D_\beta}(c)\}} \beta(\hat{\theta}), & \text{if } \theta' = \theta \\ 0, & \text{otherwise} \end{cases},$$

for any $\theta' \in \mathrm{supp}(\beta)$ and for any $\theta \in \mathrm{supp}(\beta)$ with $u(\theta) \geq \overline{\boldsymbol{p}}_{D_\beta}(c)$. Notice that by construction, $\beta \in \mathrm{co}(\{\hat{\beta}^\theta\}_{\theta \geq \overline{\boldsymbol{p}}_{D_\beta}(c)})$ and hence there exists $K^\beta \in \Delta^f(\Delta(\Theta))$ such that $\beta = \sum_{\hat{\beta}} K^\beta(\hat{\beta})$. Therefore, by splitting every β according to K^β, and by the same arguments as in the proof of Lemma 3, the resulting segmentation $\hat{s} \in \Delta^f(\Theta)$ must be such that for any $\hat{\beta} \in \mathrm{supp}(\hat{s})$, $\overline{\boldsymbol{p}}_{D_{\hat{\beta}}}(c)$ is in the interval described by $\max(\mathrm{supp}(\hat{\beta}))$. Moreover, since $\{(l(\theta), u(\theta))\}_{\theta \in \Theta}$ is disjoint, it follows that $\overline{\boldsymbol{p}}_{D_{\bar{\theta}_{\hat{\beta}}}}(c) = \overline{\boldsymbol{p}}_{D_{\hat{\beta}}}(c)$. Furthermore, since for any $\beta \in \mathrm{supp}(s)$,

$$\begin{aligned}(\overline{\boldsymbol{p}}_{D_\beta}(c) - \nu) D_\beta(\overline{\boldsymbol{p}}_\beta(c)) =& (\overline{\boldsymbol{p}}_{D_\beta}(c) - \nu) \sum_{\{\theta : u(\theta) \geq \overline{\boldsymbol{p}}_{D_\beta}(c)\}} D_\theta(\overline{\boldsymbol{p}}_\beta(c)) \beta(\theta) \\ \leq& \sum_{\{\theta : u(\theta) \geq \overline{\boldsymbol{p}}_{D_\beta}(c)\}} (\overline{\boldsymbol{p}}_{D_\theta}(c) - \nu) D_\theta(\overline{\boldsymbol{p}}_{D_\theta}(c)) \beta(\theta) \\ =& \sum_{\hat{\beta} \in \mathrm{supp}(K^\beta)} (\overline{\boldsymbol{p}}_{D_{\hat{\beta}}}(c) - \nu) D_{\hat{\beta}}(\overline{\boldsymbol{p}}_{D_{\hat{\beta}}}(c)) K^\beta(\hat{\beta}).\end{aligned}$$

As a result, since $\hat{s}(\hat{\beta}) = \sum_\beta K^\beta(\hat{\beta}) s(\beta)$, it then follows that

$$\int_{\Delta(\Theta)} (\overline{\boldsymbol{p}}_{D_\beta}(c) - \nu) D_\beta(\overline{\boldsymbol{p}}_{D_\beta}(c)) s(\mathrm{d}\beta) \leq \int_{\Delta(\Theta)} (\overline{\boldsymbol{p}}_{D_\beta}(c) - \nu) D_\beta(\overline{\boldsymbol{p}}_{D_\beta}(c)) \hat{s}(\mathrm{d}\beta).$$

Finally, since for any $\beta \in \mathrm{supp}(\hat{s})$, either $\beta(\{\theta : u(\theta) < c\}) = 1$ or $\overline{\boldsymbol{p}}_{D_\beta}(c) = \overline{\boldsymbol{p}}_{D_{\bar{\theta}_\beta}}(c)$, it must be that

$$\int_{\Delta(\Theta)} (\overline{\boldsymbol{p}}_{D_\beta}(c) - \nu) D_\beta(\overline{\boldsymbol{p}}_{D_\beta}(c)) \hat{s}(\mathrm{d}\beta) \leq \int_{\{\theta : \overline{\boldsymbol{p}}_{D_\theta}(c) \geq \nu\}} (\overline{\boldsymbol{p}}_{D_\theta}(c) - \nu) D_\theta(\overline{\boldsymbol{p}}_{D_\theta}(c)) \beta_0(\mathrm{d}\theta),$$

as desired. ∎

Lemma 18. *Suppose that D_0 is regular. For any $c \in C$ and for any $\nu \in [c, \overline{\boldsymbol{p}}_0(c)]$,*

$$D_0(\overline{\boldsymbol{p}}_0(c)) \leq \sum_{\{\theta : u(\theta) \geq \nu\}} D_\theta(\overline{\boldsymbol{p}}_{D_\theta}(c)) \beta_0(\theta) \tag{74}$$

and

$$D_0(\overline{\boldsymbol{p}}_0(c)) \geq \sum_{\{\theta: l(\theta)\geq \overline{\boldsymbol{p}}_0(c)\}} D_\theta(\overline{\boldsymbol{p}}_{D_\theta}(c))\beta_0(\theta). \tag{75}$$

Proof. Consider any $c \in C$. I first show that for any $\theta \in \Theta$ such that $l(\theta) \geq \overline{\boldsymbol{p}}_0(c)$, $\overline{\boldsymbol{p}}_{D_\theta}(c) = l(\theta)$. Indeed, since D_0 is regular, for any $\theta \in \Theta$ such that $l(\theta) \geq \overline{\boldsymbol{p}}_0(c)$ and for any $p \in (l(\theta), u(\theta)]$,

$$\begin{aligned}
&(p-c)\left[D_{\theta_p}(p)\beta_0(\theta_p) + \sum_{\{\theta': l(\theta')\geq p\}} \beta_0(\theta')\right] \\
=&(p-c)\sum_{\{\theta': u(\theta')\geq p\}} D_{\theta'}(p)\beta_0(\theta') \\
=&(p-c)D_0(p) \\
\leq&(l(\theta)-c)D_0(l(\theta)) \\
=&(l(\theta)-c)\left[\sum_{\{\theta': u(\theta')\geq l(\theta)\}} D_{\theta'}(l(\theta))\beta_0(\theta')\right] \\
=&(l(\theta)-c)\left[D_\theta(l(\theta))\beta_0(\theta) + \sum_{\{\theta': l(\theta')\geq l(\theta)\}} \beta_0(\theta')\right].
\end{aligned}$$

As such, since $p \in (l(\theta), u(\theta)]$ and $u(\theta_p) = u(\theta)$, it must be that

$$(p-c)D_\theta(p) < (l(\theta)-c)D_\theta(l(\theta)),$$

which then implies that $\overline{\boldsymbol{p}}_{D_\theta}(c) = l(\theta)$.

Now, I show that $\overline{\boldsymbol{p}}_0(c) \geq \widehat{\boldsymbol{p}}_0(c) := \overline{\boldsymbol{p}}_{D_{\theta_{\overline{\boldsymbol{p}}_0(c)}}}(c)$. Indeed, by definitions,

$$\begin{aligned}
=&(\widehat{\boldsymbol{p}}_0(c)-c)\left[D_{\theta_{\overline{\boldsymbol{p}}_0(c)}}(\widehat{\boldsymbol{p}}_0(c))\beta_0(\theta_{\overline{\boldsymbol{p}}_0(c)}) + \sum_{\{\theta': l(\theta')\geq \widehat{\boldsymbol{p}}_0(c)\}} \beta_0(\theta')\right] \\
=&(\widehat{\boldsymbol{p}}_0(c)-c)D_0(\widehat{\boldsymbol{p}}_0(c)) \\
\leq&(\overline{\boldsymbol{p}}_0(c)-c)D_0(\overline{\boldsymbol{p}}_0(c)) \\
=&(\overline{\boldsymbol{p}}_0(c)-c)\left[D_{\theta_{\overline{\boldsymbol{p}}_0(c)}}(\overline{\boldsymbol{p}}_0(c)) + \sum_{\{\theta': l(\theta')\geq \overline{\boldsymbol{p}}_0(c)\}} \beta_0(\theta')\right],
\end{aligned}$$

and

$$(\overline{\boldsymbol{p}}_0(c)-c)D_{\theta_{\overline{\boldsymbol{p}}_0(c)}}(\overline{\boldsymbol{p}}_0(c)) \leq (\widehat{\boldsymbol{p}}_0(c)-c)D_{\theta_{\overline{\boldsymbol{p}}(c)}}(\widehat{\boldsymbol{p}}_0(c)).$$

As a result, it must be that $\widehat{\boldsymbol{p}}_0(c) \leq \overline{\boldsymbol{p}}_0(c)$.

Consequently,

$$\begin{aligned}
\sum_{\{\theta: l(\theta)\geq \overline{\boldsymbol{p}}_0(c)\}} D_\theta(\overline{\boldsymbol{p}}_\theta(c))\beta_0(\theta) =& \sum_{\{\theta: l(\theta)\geq \overline{\boldsymbol{p}}_0(c)\}} \beta_0(\theta) \\
\leq& \sum_{\{\theta: l(\theta)\geq \overline{\boldsymbol{p}}_0(c)\}} \beta_0(\theta) + D_{\theta_{\overline{\boldsymbol{p}}_0(c)}}(\overline{\boldsymbol{p}}_0(c))\beta_0(\theta_{\overline{\boldsymbol{p}}_0(c)}) \\
\leq& D_0(\overline{\boldsymbol{p}}_0(c)),
\end{aligned}$$

which proves (75). On the other hand, for any $\nu \in [c, \overline{\boldsymbol{p}}_0(c)]$

$$
\begin{aligned}
&\sum_{\{\theta: u(\theta)\geq \nu\}} D_\theta(\overline{\boldsymbol{p}}_{D_\theta}(c))\beta_0(\theta)\\
&= \sum_{\{\theta: \nu\leq u(\theta)<\overline{\boldsymbol{p}}_0(c)\}} D_\theta(\overline{\boldsymbol{p}}_{D_\theta}(c))\beta_0(\theta) + \sum_{\{\theta: u(\theta)\geq \overline{\boldsymbol{p}}_0(c)\}} D_\theta(\overline{\boldsymbol{p}}_{D_\theta}(c))\beta_0(\theta)\\
&\geq \sum_{\{\theta: u(\theta)\geq \overline{\boldsymbol{p}}_0(c)\}} D_\theta(\overline{\boldsymbol{p}}_{D_\theta}(c))\beta_0(\theta)\\
=&D_{\theta_{\overline{\boldsymbol{p}}_0(c)}}(\widehat{\boldsymbol{p}}_0(c)) + \sum_{\{\theta': l(\theta')\geq \overline{\boldsymbol{p}}_0(c)\}} D_{\theta'}(l(\theta'))\beta_0(\theta')\\
\geq&D_{\theta_{\overline{\boldsymbol{p}}_0(c)}}(\overline{\boldsymbol{p}}_0(c)) + \sum_{\{\theta': l(\theta')\geq \overline{\boldsymbol{p}}_0(c)\}} \beta_0(\theta')\\
=&D_0(\overline{\boldsymbol{p}}_0(c)),
\end{aligned}
$$

which proves (74) ■

Proof of Theorem 9. To prove Theorem 9, first notice that Lemma 1 still applies and hence the data broker's maximization problem can be written as

$$
\begin{aligned}
&\max_\sigma \int_C \left(\int_{\Delta(\Theta)} (\overline{\boldsymbol{p}}_{D_\beta}(c)-\phi_G(c))D_\beta(\overline{\boldsymbol{p}}_{D_\beta}(c))\sigma(\mathrm{d}\beta|c)\right) G(\mathrm{d}c)\\
&\text{s.t. } \int_c^{c'} \big(D_\beta(\overline{\boldsymbol{p}}_\beta(z))(\sigma(\mathrm{d}\beta|z)-\sigma(\mathrm{d}\beta|c'))\big)\,\mathrm{d}z \geq 0,\ \forall c,c'\in C\\
&\qquad \bar{\pi} + \int_c^{\bar{c}} \left(\int_{\Delta(\Theta)} D_\beta(\overline{\boldsymbol{p}}_\beta(z))\sigma(\mathrm{d}\beta|z)\right)\mathrm{d}z \geq \bar{\pi} + \int_c^{\bar{c}} D_0(\overline{\boldsymbol{p}}_0(z))\,\mathrm{d}z,\ \forall c\in C,
\end{aligned}
\tag{76}
$$

where the maximum is taken over all $\sigma : C \to \Delta(\Delta(\Theta))$ such that $\sigma(c)$ is a segmentation for all $c \in C$.

Consider first a relaxed problem of (76) where the first constraint is relaxed to $\boldsymbol{D}_\sigma \in [0,1]^C$ being nonincreasing, where

$$
\boldsymbol{D}_\sigma(c) := \int_{\Delta(\Theta)} D_\beta(\overline{\boldsymbol{p}}_{D_\beta}(c))\sigma(\mathrm{d}\beta|c),
$$

for all $c \in C$. By the same duality argument as in the proof of Lemma 16, it suffices to find a feasible σ^* and a Borel measure μ^* on C such that

$$
\begin{aligned}
\sigma^* \in \operatorname*{argmax}_{\sigma\in\Sigma}\Bigg[&\int_C \left(\int_{\Delta(\Theta)} (\overline{\boldsymbol{p}}_{D_\beta}(c)-\phi_G(c))D_\beta(\overline{\boldsymbol{p}}_{D_\beta}(c))\sigma(\mathrm{d}\beta|c)\right) G(\mathrm{d}c)\\
&+\int_C \left(\int_c^{\bar{c}} \left(\int_{\Delta(\Theta)} D_\beta(\overline{\boldsymbol{p}}_\beta(z))\sigma(\mathrm{d}\beta|z) - D_0(\overline{\boldsymbol{p}}_0(z))\right)\mathrm{d}z\right)\mu^*(\mathrm{d}c)\Bigg],
\end{aligned}
$$

where Σ is the collection of segmentation schemes such that $\boldsymbol{D}_\sigma$ is nonincreasing, and that

$$
\int_C \left(\int_c^{\bar{c}} \left(\int_{\Delta(\Theta)} D_\beta(\overline{\boldsymbol{p}}_{D_\beta}(z))\sigma^*(\mathrm{d}\beta|z) - D_0(\overline{\boldsymbol{p}}_0(z))\right)\mathrm{d}z\right)\mu^*(\mathrm{d}c) = 0.
$$

To this end, let M^* be defined as

$$
M^*(c) := \lim_{c'\downarrow c} g(c)(\phi_G(c) - \widehat{\boldsymbol{p}}_0(c))^+.
$$

Since $c \mapsto g(c)(\phi_G(c) - \widehat{\boldsymbol{p}}_0(c))^+$ in nondecreasing, M^* is nondecreasing and right-continuous and hence induced a Borel measure μ^* with $\mathrm{supp}(\mu^*) = [c^*, \overline{c}]$ for some $c^* \leq \overline{c}$. Then, by the same arguments as in the proof of Lemma 16 (in particular, the definition of $\widehat{\varphi}_G$, (37) and (38)),

$$\max_{\sigma\in\Sigma}\Bigg[\int_C \left(\int_{\Delta(\Theta)} (\overline{\boldsymbol{p}}_{D_\beta}(c) - \phi_G(c)) D_\beta(\overline{\boldsymbol{p}}_{D_\beta}(c))\sigma(\mathrm{d}\beta|c)\right) G(\mathrm{d}c) \\ + \int_C \left(\int_c^{\overline{c}} \left(\int_{\Delta(\Theta)} D_\beta(\overline{\boldsymbol{p}}_\beta(z))\sigma(\mathrm{d}\beta|z) - D_0(\overline{\boldsymbol{p}}_0(z))\right) \mathrm{d}z\right) \mu^*(\mathrm{d}c)\Bigg]$$

is equivalent to

$$\max_{\sigma\in\Sigma}\int_C \left(\int_{\Delta(\Theta)} (\overline{\boldsymbol{p}}_{D_\beta}(c) - \widehat{\varphi}_G(c)) D_\beta(\overline{\boldsymbol{p}}_{D_\beta}(c))\sigma(\mathrm{d}\beta|c)\right) G(\mathrm{d}c). \tag{77}$$

To solve (77), notice that for any $c \in [\underline{c}, c^*)$,

$$\sum_{\{\theta: u(\theta)\geq\widehat{\varphi}_G(c)\}} D_\theta(\overline{\boldsymbol{p}}_{D_\theta}(c)) > D_0(\overline{\boldsymbol{p}}_0(c)),$$

which is due to $\widehat{\varphi}_G(c) = \varphi_G(c) \leq \widehat{\boldsymbol{p}}_0(c) \leq \overline{\boldsymbol{p}}_0(c)$ and (75). On the other hand, for any $c \in (c^*, \overline{c}]$, there exists a unique $\lambda(c)$ such that

$$\lambda(c) D_{\theta_{\widehat{\varphi}_G(c)}}(\widehat{\boldsymbol{p}}_0(c)) + \sum_{\{\theta: l(\theta)\geq\widehat{\varphi}_G(c)\}} D_\theta(\overline{\boldsymbol{p}}_{D_\theta}(c)) = D_0(\overline{\boldsymbol{p}}_0(c)),$$

which is due to the fact that $\widehat{\varphi}_G(c) = \widehat{\boldsymbol{p}}_0(c)$ for all $c \in (c^*, \overline{c}]$ and (74). Furthermore, Since D_0 is regular, for any $\theta \in \Theta$ such that $u(\theta) \geq \widehat{\varphi}_G(c)$ and for any $p \leq l(\theta_{\widehat{\varphi}_G(c)})$,

$$\begin{aligned}
(p-c)D_{\beta^\theta_{\widehat{\varphi}_G(c)}}(p) = & \sum_{\{\theta': u(\theta')\geq u(\theta_p)\}} (p-c) D_{\theta'}(p)\beta^\theta_{\widehat{\varphi}_G(c)}(\theta') \\
=& (p-c)D_0(p) \\
\leq& (l(\theta_{\widehat{\varphi}_G(c)}) - c) D_0(l(\theta_{\widehat{\varphi}_G(c)})) \\
\leq& (l(\theta) - c) D_0(l(\theta_{\widehat{\varphi}_G(c)})) \\
=& (l(\theta) - c) \sum_{\{\theta': u(\theta')\geq\widehat{\varphi}_G(c)\}} \beta^\theta_{\widehat{\varphi}_G(c)}(\theta') \\
=& (l(\theta) - c) D_{\beta^\theta_{\widehat{\varphi}_G(c)}}(l(\theta)) \\
=& (\overline{\boldsymbol{p}}_{D_\theta}(c) - c) D_{\beta^\theta_{\widehat{\varphi}_G(c)}}(\overline{\boldsymbol{p}}_{D_\theta}(c)),
\end{aligned} \tag{78}$$

where $\beta^\theta_{\widehat{\varphi}_G(c)}$ is defined in (11). In addition, by the same construction as in the proof of Lemma 3, for any $c \in (c^*, \overline{c}]$, there exists a segmentation $\tilde{\sigma}(c) \in \Delta^f(\Delta(\Theta))$ such that $\mathrm{supp}(\tilde{\sigma}(c)) = \{\tilde{\beta}^\theta_{\widehat{\boldsymbol{p}}_0(c)} : l(\theta) \geq \widehat{\boldsymbol{p}}_0(c)\}$, with $\tilde{\beta}^\theta_{\widehat{\boldsymbol{p}}_0(c)}$ satisfying (12) and (13) and that

$$(p-c)D_{\tilde{\beta}^\theta_{\widehat{\boldsymbol{p}}_0(c)}}(p) \leq (l(\theta) - c)D_\theta(l(\theta)) = (\overline{\boldsymbol{p}}_{D_\theta}(c) - c)D_\theta(\overline{\boldsymbol{p}}_{D_\theta}(c)) \tag{79}$$

for all $\theta \in \Theta$ such that $l(\theta) \geq \overline{\boldsymbol{p}}_0(c)$, as well as

$$\overline{\boldsymbol{p}}_{D_{\tilde{\beta}^\theta_{\widehat{\boldsymbol{p}}_0(c)}}}(z) \geq \overline{\boldsymbol{p}}_{D_0}(z) \geq \widehat{\boldsymbol{p}}_0(z) \tag{80}$$

for all $z \in [\underline{c}, c]$ and for all $\theta \in \Theta$ such that $l(\theta) \geq \overline{\boldsymbol{p}}_0(c)$.

Now define σ^* as follows.

$$\sigma^*(c) := \begin{cases} \sigma_1(c), & \text{if } c \in [\underline{c}, c^*] \\ \sigma_2(c), & \text{if } c \in (c^*, \overline{c}] \end{cases},$$

where

$$\sigma_1(\beta^\theta_{\varphi_G(c)}|c) := \frac{\beta_0(\theta)}{\sum_{\{\theta': u(\theta') \geq \varphi_G(c)\}} \beta_0(\theta')}$$

for all $c \in [\underline{c}, c^*]$ and for all $\theta \in \Theta$ such that $u(\theta) \geq \varphi_G(c)$, whereas

$$\sigma_2(\beta|c) := \begin{cases} \lambda(c)\frac{\beta_0(\theta)}{\sum_{\{\theta': u(\theta') \geq \widehat{\boldsymbol{p}}_0(c)\}} \beta_0(\theta')}, & \text{if } \beta = \beta^\theta_{\widehat{\boldsymbol{p}}_0(c)},\, u(\theta) \geq \widehat{\boldsymbol{p}}_0(c) \\ (1-\lambda(c))\tilde{\sigma}(\tilde{\beta}^\theta_{\widehat{\boldsymbol{p}}_0(c)}|c), & \text{if } \beta = \tilde{\beta}^\theta_{\widehat{\boldsymbol{p}}_0(c)},\, l(\theta) \geq \widehat{\boldsymbol{p}}_0(c) \\ 0, & \text{otherwise} \end{cases},$$

for all $c \in (c^*, \overline{c}]$. It then follows that, by (78) and (79),

$$\begin{aligned} &\int_C \left(\int_{\Delta(\Theta)} (\overline{\boldsymbol{p}}_{D_\beta}(c) - \widehat{\varphi}_G(c)) D_\beta(\overline{\boldsymbol{p}}_{D_\beta}(c)) \sigma^*(\mathrm{d}\beta|c) \right) G(\mathrm{d}c) \\ =& \int_C \left(\sum_{\{\theta: \overline{\boldsymbol{p}}_{D_\theta}(c) \geq \widehat{\varphi}_G(c)\}} (\overline{\boldsymbol{p}}_{D_\theta}(c) - \widehat{\varphi}_G(c)) D_\theta(\overline{\boldsymbol{p}}_{D_\theta}(c)) \beta_0(\theta) \right) G(\mathrm{d}c), \end{aligned}$$

which, together with Lemma 17, implies that σ^* is a solution of (77).

Furthermore, for any $c > c^*$, by the definition of $\sigma_2(c)$ and $\lambda(c)$, by (78) and (79), and by the fact that $\widehat{\varphi}_G(c) = \widehat{\boldsymbol{p}}_0(c)$,

$$\int_{\Delta(\Theta)} D_\beta(\overline{\boldsymbol{p}}_\beta(c)) \sigma^*(\mathrm{d}\beta|c) = D_0(\overline{\boldsymbol{p}}_0(c)).$$

Therefore,

$$\int_C \left(\int_c^{\overline{c}} \left(\int_{\Delta(\Theta)} D_\beta(\overline{\boldsymbol{p}}_\beta(z)) \sigma^*(\mathrm{d}\beta|z) - D_0(\overline{\boldsymbol{p}}_0(z)) \right) \mathrm{d}z \right) \mu^*(\mathrm{d}c) = 0.$$

Finally, by definition of $\widehat{\varphi}_G$ and by Lemma 18,

$$\int_{\Delta(\Theta)} D_\beta(\overline{\boldsymbol{p}}_\beta(c)) \sigma^*(\mathrm{d}\beta|c) \geq D_0(\overline{\boldsymbol{p}}_0(c))$$

for all $c \in [\underline{c}, c^*)$. Together with monotonicity of $\widehat{\varphi}_G$, $\sigma^* \in \Sigma$ and is a solution of the relaxed problem of (76).

It then suffices to show that σ^* is implementable. Notice that for any $c \in C$ and for any $z \in [\underline{c}, c]$ and for any $\beta^\theta_{\widehat{\varphi}_G(c)} \in \mathrm{supp}(\sigma^*(c))$, if

$$\boldsymbol{P}_{D_{\beta^\theta_{\widehat{\varphi}_G(c)}}}(z) \cap \mathrm{supp}(D_\theta) = \emptyset,$$

then it must be that

$$\begin{aligned} (p-z)D_0 =& (p-z) D_{\beta^\theta_{\widehat{\varphi}_G(c)}}(p) \\ \leq& (\overline{\boldsymbol{p}}_{D_{\beta^\theta_{\widehat{\varphi}_G(c)}}}(z) - z) D_{\beta^\theta_{\widehat{\varphi}_G(c)}}(\overline{\boldsymbol{p}}_{D_{\beta^\theta_{\widehat{\varphi}_G(c)}}}(z)) \\ =& (\overline{\boldsymbol{p}}_{D_{\beta^\theta_{\widehat{\varphi}_G(c)}}}(z) - z) D_0(\overline{\boldsymbol{p}}_{D_{\beta^\theta_{\widehat{\varphi}_G(c)}}}(z)), \end{aligned}$$

for all $p \leq \overline{\boldsymbol{p}}_{D_{\beta^{\theta}_{\widehat{\varphi}_G(c)}}}(z)$. Therefore,

$$\overline{\boldsymbol{p}}_{D_{\beta^{\theta}_{\widehat{\varphi}_G(c)}}}(z) \geq \overline{\boldsymbol{p}}_0(z) \geq \widehat{\boldsymbol{p}}_0(z) \geq \widehat{\varphi}_G(z),$$

for all $z \in [\underline{c}, c]$. Together with (80), by the same argument as the proof of Lemma 2, σ^* is indeed implementable. This completes the proof. ■

F.3 Proof of Theorem 10

Before proving Theorem 10. I first introduced the counterpart of Lemma 1 when targeting is available. The proof of this result is entirely analogous to the proof of Lemma 1 and therefore omitted.

Lemma 19. *A mechanism (σ, τ, q) is incentive compatible if and only if there exists constants $\{\overline{\tau}_j\}_{j\in\mathcal{J}} \subset \mathbb{R}$ such that for all $j \in \mathcal{J}$ and for all $c_j, c_j' \in C_j$,*

1.
$$\begin{aligned}
&\mathbb{E}_{c_{-j}}[\tau_j(c)] \\
=&\mathbb{E}_{c_{-j}}\left[\sum_{i\in\mathcal{I}}\int_{\mathcal{D}} \pi_D(c_j)\sigma_{ij}(\mathrm{d}D|c)q_{ij}(c)\right] \\
&- \int_{c_j}^{\overline{c}_j} \mathbb{E}_{c_{-j}}\left[\sum_{i\in\mathcal{I}}\int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D^j(z))\sigma_{ij}(\mathrm{d}D|z, c_{-j})q_{ij}(z, c_{-j})\right]\mathrm{d}z - \overline{\tau}_j,
\end{aligned}$$

2.
$$\begin{aligned}
&\int_{c_j}^{c_j'}\left(\mathbb{E}_{c_{-j}}\left[\sum_{i\in\mathcal{I}}\int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D^j(z))\sigma_{ij}(\mathrm{d}D|z, c_{-j})q_{ij}(z, c_{-j})\right]\right)\mathrm{d}z \\
&- \int_{c_j}^{c_j'}\left(\mathbb{E}_{c_{-j}}\left[\sum_{i\in\mathcal{I}}\int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D^j(z))\sigma_{ij}(\mathrm{d}D|c_j', c_{-j})\right]q_{ij}(c_j', c_{-j})\right)\mathrm{d}z \geq 0
\end{aligned}$$

With the characterization given by Lemma 19, the data broker's expected revenue can again be written as

$$\int_C \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}}\left(\int_{\mathcal{D}}(\overline{\boldsymbol{p}}_D^j(c_j) - \phi_{G_j}(c_j))D(\overline{\boldsymbol{p}}_D^j(c_j))\sigma_{ij}(\mathrm{d}D|c)q_{ij}(c)\right)G(\mathrm{d}c) - \sum_{j\in\mathcal{J}}\bar{\pi}_j,$$

where $\bar{\pi}_j := \pi_{D_0^j}(\overline{c}_j)$. Using this observation,

Proof of Theorem 10. Existence of solutions is ensured by compactness of the feasible set and continuity of the objective function, which rely on Lemma 8, Tychonoff's theorem, and the Lebesgue dominate convergence theorem.

Now consider any mechanism (σ, τ, q). Suppose that the consumers retain positive surplus. That is

$$\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}}\int_C\int_{\mathcal{D}}(v - \overline{\boldsymbol{p}}_D(c_j))D(\mathrm{d}v)\sigma_{ij}(\mathrm{d}D|c)q_{ij}(c)G(\mathrm{d}c) > 0.$$

It then suffices to show that there exists an incentive feasible mechanism $(\hat{\sigma}, \hat{\tau}, \hat{q})$ that strictly improves the data broker's revenue.

Notice that under (σ, τ, q), the data broker's revenue is

$$\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}}\int_C\int_{\mathcal{D}}(\overline{\boldsymbol{p}}_D(c_j)-\phi_{G_j}(c_j))D(\overline{\boldsymbol{p}}_D(c_j))\sigma_{ij}(\mathrm{d}D|c)q_{ij}(c)G(\mathrm{d}c)-\sum_{j\in\mathcal{J}}U_j(\overline{c}_j), \tag{81}$$

where U_j is the indirect utility of producer j. On the other hand, notice that for any $i \in \mathcal{I}$, $j \in \mathcal{J}$, since the segmentation scheme $\sigma_{ij} \in \mathcal{S}^C_{D_0^{ij}}$ is measurable, the mapping $\overline{\sigma}_{ij} : C_j \to \mathcal{D}$, defined as

$$\overline{\sigma}_{ij}(c_j) := \mathbb{E}_{c_{-j}}[\sigma_{ij}(c_j, c_{-j})],\, \forall c_j \in C_j$$

is also measurable and thus is also in $\mathcal{S}^C_{D_0^{ij}}$. As a result, as shown in the proof of Theorem S1 in the Supplemental Material, for any $j \in \mathcal{J}$ and any $i \in \mathcal{I}$, there exists a measurable function $\tilde{\sigma}_{ij} : C_j \to \mathcal{D}$ such that

$$\int_{\mathcal{D}} D\tilde{\sigma}_{ij}(\mathrm{d}D|c_j) = D_0^{ij},\, \forall c_j \in C_j, \tag{82}$$

and that

$$\begin{aligned}
&\int_{C_j}\int_{\mathcal{D}}(\overline{\boldsymbol{p}}_D(c_j)-\phi_{G_j}(c_j))D(\overline{\boldsymbol{p}}_D(c_j))\tilde{\sigma}_{ij}(\mathrm{d}D|c_j)G_j(\mathrm{d}c)\\
\geq&\int_{C_j}\int_{\mathcal{D}}(\overline{\boldsymbol{p}}_D(c_j)-\phi_{G_j}(c_j))D(\overline{\boldsymbol{p}}_D(c_j))\overline{\sigma}_{ij}(\mathrm{d}D|c_j)G_j(\mathrm{d}c_j)\\
&+\int_{C_j}\int_{\mathcal{D}}\int_{\{v\geq\overline{\boldsymbol{p}}_D(c_j)\}}(v-\overline{\boldsymbol{p}}_D(c_j))D(\mathrm{d}v)\overline{\sigma}_{ij}(\mathrm{d}D|c_j)G_j(\mathrm{d}c_j)\\
\geq&\int_{C_j}\int_{\mathcal{D}}(\overline{\boldsymbol{p}}_D(c_j)-\phi_{G_j}(c_j))D(\overline{\boldsymbol{p}}_D(c_j))\overline{\sigma}_{ij}(\mathrm{d}D|c)G_j(\mathrm{d}c_j).
\end{aligned} \tag{83}$$

By (81), there exists $i^* \in \mathcal{I}$ and $j^* \in \mathcal{J}$ such that

$$\int_{C_j}\int_{\mathcal{D}}\int_{\{v\geq\overline{\boldsymbol{p}}_D(c_j)\}}(v-\overline{\boldsymbol{p}}_D(c_j))D(\mathrm{d}v)\overline{\sigma}_{ij}(\mathrm{d}D|c_j)\overline{q}_{ij}(c_j)G_j(\mathrm{d}c_j)>0,$$

where $\overline{q}_{ij}(c_j) := \mathbb{E}_{c_{-j}}[q_{ij}(c_j, c_{-j})]$ for all $i \in \mathcal{I}$, $j \in \mathcal{J}$, $c_j \in C_j$. As such, since $q_{i^*j^*} \in [0, 1]$, we must have

$$\int_{C_j^*}\int_{\mathcal{D}}\int_{\{v\geq\overline{\boldsymbol{p}}_D(c_j^*)\}}(v-\overline{\boldsymbol{p}}_D(c_j^*))D(\mathrm{d}v)\overline{\sigma}_{i^*j^*}(\mathrm{d}D|c_{j^*})\overline{q}_{i^*j^*}(c_j^*)G(\mathrm{d}c_j^*)>0$$

and hence the last inequality in (83) must be strict inequality for some $i \in \mathcal{I}$, $j \in \mathcal{J}$. Therefore, if $\hat{\sigma} \in \mathcal{D}^C$ is defined as

$$\hat{\sigma}_{ij}(c) := \tilde{\sigma}_{ij}(c_j),\, \forall i \in \mathcal{I}, j \in \mathcal{J}, c \in C,$$

then

$$\begin{aligned}
&\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}}\int_C\int_{\mathcal{D}}(\overline{\boldsymbol{p}}_D(c_j)-\phi_{G_j}(c_j))D(\overline{\boldsymbol{p}}_D(c_j))\hat{\sigma}_{ij}(\mathrm{d}D|c)q_{ij}(c)G(\mathrm{d}c)\\
>&\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}}\int_C\int_{\mathcal{D}}(\overline{\boldsymbol{p}}_D(c_j)-\phi_{G_j}(c_j))D(\overline{\boldsymbol{p}}_D(c_j))\sigma_{ij}(\mathrm{d}D|c)q_{ij}(c)G(\mathrm{d}c).
\end{aligned} \tag{84}$$

On the other hand, as shown in the proof of Theorem S1 in the Supplemental Material, such $\{\tilde{\sigma}_{ij}\}$ are such that for any $i \in \mathcal{I}$ and $j \in \mathcal{J}$, for Lebesgue-almost all $c_j \in C_j$,

$$\int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(c_j))\tilde{\sigma}_{ij}(\mathrm{d}D|c_j) = \int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(c_j))\overline{\sigma}_{ij}(\mathrm{d}D|c_j). \tag{85}$$

Moreover, for all $c_j, c_j' \in C_j$ with $c_j' > c_j$,

$$\int_{c_j}^{c_j'} \left(\int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(z))\tilde{\sigma}_{ij}(\mathrm{d}D|z)\right) \mathrm{d}z \geq \int_{c_j}^{c_j'} \left(\int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(z))\tilde{\sigma}_{ij}(\mathrm{d}D|c_j)\right) \mathrm{d}z, \tag{86}$$

and,

$$\int_{c_j}^{c_j'} \left(\int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(z))\tilde{\sigma}_{ij}(\mathrm{d}D|z)\right) \leq \int_{c_j}^{c_j'} \left(\int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(z))\tilde{\sigma}_{ij}(\mathrm{d}D|c_j')\right). \tag{87}$$

These imply that, as $\overline{q}_{ij}(c_j) \in [0,1]$ for all $c_j \in C_j$, for all $j \in \mathcal{J}$,

$$\begin{aligned}
&\int_{c_j}^{c_j'} \left(\mathbb{E}_{c_{-j}}\left[\sum_{i\in\mathcal{I}} \int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(z))\hat{\sigma}_{ij}(\mathrm{d}D|z, c_{-j})q_{ij}(z, c_{-j})\right]\right) \mathrm{d}z \\
&\quad - \int_{c_j}^{c_j'} \left(\mathbb{E}_{c_{-j}}\left[\sum_{i\in\mathcal{I}} \int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(z))\hat{\sigma}_{ij}(\mathrm{d}D|c_j', c_{-j})q_{ij}(c_j', c_{-j})\right]\right) \mathrm{d}z \\
=&\int_{c_j}^{c_j'} \left(\sum_{i\in\mathcal{I}} \int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(z))\tilde{\sigma}_{ij}(\mathrm{d}D|z)\overline{q}_{ij}(z)\right) \mathrm{d}z \\
&\quad - \int_{c_j}^{c_j'} \left(\sum_{i\in\mathcal{I}} \int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(z))\tilde{\sigma}_{ij}(\mathrm{d}D|c_j')\overline{q}_{ij}(c_j')\right) \mathrm{d}z \\
\geq&\int_{c_j}^{c_j'} \left(\sum_{i\in\mathcal{I}} \int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(z))\overline{\sigma}_{ij}(\mathrm{d}D|z)\overline{q}_{ij}(z)\right) \mathrm{d}z \\
&\quad - \int_{c_j}^{c_j'} \left(\sum_{i\in\mathcal{I}} \int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(z))\overline{\sigma}_{ij}(\mathrm{d}D|c_j')\overline{q}_{ij}(c_j')\right) \mathrm{d}z \\
=&\int_{c_j}^{c_j'} \left(\mathbb{E}_{c_{-j}}\left[\sum_{i\in\mathcal{I}} \int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(z))\sigma_{ij}(\mathrm{d}D|z, c_{-j})q_{ij}(z, c_{-j})\right]\right) \mathrm{d}z \\
&\quad - \int_{c_j}^{c_j'} \left(\mathbb{E}_{c_{-j}}\left[\sum_{i\in\mathcal{I}} \int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(z))\sigma_{ij}(\mathrm{d}D|c_j', c_{-j})q_{ij}(c_j', c_{-j})\right]\right) \mathrm{d}z \\
\geq&0,
\end{aligned} \tag{88}$$

for all $j \in \mathcal{J}$, $c_j, c_j' \in C_j$, where the first equality follows from the definitions of $\hat{\sigma}_{ij}$ and $\overline{q}_{ij}$, the first inequality follows from (85), (86) and (87), the second equality follows from the definition of $\overline{\sigma}_{ij}$ and the last inequality follows from the incentive compatibility of (σ, τ, q). Furthermore, for any $j \in \mathcal{J}$, $c_j \in C_j$,

$$\begin{aligned}
&\int_{c_j}^{\bar{c}_j} \left(\mathbb{E}_{c_j} \left[\sum_{i \in \mathcal{I}} \int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(z)) \hat{\sigma}_{ij}(\mathrm{d}D|z, c_{-j}) q(z, c_{-j}) \right] \right) \mathrm{d}z \\
&= \int_{c_j}^{\bar{c}_j} \left(\sum_{i \in \mathcal{I}} \int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(z)) \tilde{\sigma}_{ij}(\mathrm{d}D|z) \overline{q}_{ij}(z) \right) \mathrm{d}z \\
&= \int_{c_j}^{\bar{c}_j} \left(\sum_{i \in \mathcal{I}} \int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(z)) \overline{\sigma}_{ij}(\mathrm{d}D|z) \overline{q}_{ij}(z) \right) \mathrm{d}z \qquad (89) \\
&= \int_{c_j}^{\bar{c}_j} \left(\mathbb{E}_{c_j} \left[\sum_{i \in \mathcal{I}} \int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(z)) \sigma_{ij}(\mathrm{d}D|z, c_{-j}) q(z, c_{-j}) \right] \right) \mathrm{d}z \\
&\geq \pi_j(c_j, m_j^*) - \bar{\pi}_j,
\end{aligned}$$

where the first equality follows from the definition of $\hat{\sigma}_{ij}$ and $\overline{q}_{ij}$, the second equality follows from (85), the third equality follows from the definition of $\overline{\sigma}_{ij}$ and the last inequality follows from the individual rationality of (σ, τ, q) and Lemma 19.

Together, by (88), (89), (82) and Lemma 19, there exist transfers $\{\hat{\tau}_j\}$ such that $(\hat{\sigma}, \hat{\tau}, q)$ is an incentive compatible and individually rational mechanism. Moreover, by (84), this mechanism improves the data broker's revenue. As such, (σ, τ, q) cannot be optimal. This completes the proof. ■

F.4 Proof of Theorem 11

Proof of Theorem 11. First notice that by Lemma 19, for any incentive feasible mechanism (σ, τ, q), the data broker's expected revenue is at most

$$\sum_{j \in \mathcal{J}} \left[\int_{C_j} \left(\sum_{i \in \mathcal{I}} \int_{\mathcal{D}} (\overline{\boldsymbol{p}}_D(c_j) - \phi_{G_j}(c_j)) D(\overline{\boldsymbol{p}}_D(c_j)) \overline{\sigma}_{ij}(\mathrm{d}D|c_j) \overline{q}_{ij}(c) \right) G_j(\mathrm{d}c_j) - \bar{\pi}_j \right],$$

where for any $i \in \mathcal{I}$, $j \in \mathcal{J}$, $c_j \in C_j$,

$$\overline{\sigma}_{ij}(c_j) := \mathbb{E}_{c_j}[\sigma_{ij}(c)], \quad \overline{q}_{ij}(c_j) := \mathbb{E}_{c_{-j}}[q_{ij}(c)].$$

Furthermore, notice that for any $i \in \mathcal{I}$, $j \in \mathcal{J}$ and any $c_j \in C_j$,

$$\begin{aligned}
\int_{\mathcal{D}} (\overline{\boldsymbol{p}}_D(c_j) - \phi_{G_j}(c_j)) D(\overline{\boldsymbol{p}}_D(c_j)) \overline{\sigma}_{ij}(\mathrm{d}D|c_j) &\leq \int_{\mathcal{D}} \max_{p \in \mathbb{R}_+} (p - \phi_{G_j}(c_j)) D(p) \overline{\sigma}_{ij}(\mathrm{d}D|c_j) \\
&\leq \int_V (v - \phi_{G_j}(c_j))^+ D_0^{ij}(\mathrm{d}v).
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\sum_{j\in\mathcal{J}}\left[\int_{C_j}\left(\sum_{i\in\mathcal{I}}\int_{\mathcal{D}}(\overline{\boldsymbol{p}}_D(c_j)-\phi_{G_j}(c_j))D(\overline{\boldsymbol{p}}_D(c_j))\overline{\sigma}_{ij}(\mathrm{d}D|c_j)\overline{q}_{ij}(c)\right)G_j(\mathrm{d}c_j)\right]\\
\leq&\sum_{j\in\mathcal{J}}\left[\int_{C_j}\left(\sum_{i\in\mathcal{I}}\int_{\mathcal{D}}\max_{p\in\mathbb{R}_+}(p-\phi_{G_j}(c_j))D(p)\overline{\sigma}_{ij}(\mathrm{d}D|c_j)\overline{q}_{ij}(c)\right)G_j(\mathrm{d}c_j)\right] \qquad (90)\\
\leq&\sum_{j\in\mathcal{J}}\left[\int_{C_j}\left(\int_V(v-\phi_{G_j}(c_j))^+D_0^{ij}(\mathrm{d}v)\right)\overline{q}_{ij}(c_j)G_j(\mathrm{d}c_j)\right]\\
=&\int_C\sum_{j\in\mathcal{J}}\left(\sum_{i\in\mathcal{I}}\left(\int_V(v-\phi_{G_j}(c_j))^+D_0^{ij}(\mathrm{d}v)\right)q_{ij}(c)\right)G(\mathrm{d}c).
\end{aligned}$$

Since $\sum_{i\in\mathcal{I}} q_{ij}\leq 1$ for all $j\in\mathcal{J}$, the data broker's revenue is bounded from above by

$$\bar{R}:=\sum_{j\in\mathcal{J}}\int_C\left[\max_{i\in\mathcal{I}}\left(\int_V(v-\phi_{G_j}(c_j))^+D_0^{ij}(\mathrm{d}v)\right)G_j(\mathrm{d}c_j)\right]-\sum_{j\in\mathcal{J}}\bar{\pi}_j.$$

I first show that there exists an incentive feasible mechanism that attains the upper bound $\bar{R}$. To see this, notice that for any $i\in\mathcal{I}$, $j\in\mathcal{J}$, since $c_j\leq\phi_{G_j}(c_j)\leq\overline{\boldsymbol{p}}_{D_0^{ij}}(c_j)$ for all $c_j\in C_j$, as shown in the proof of Theorem 1, there exists a segmentation scheme $\sigma^*_{ij}:C_j\to\mathcal{D}$ such that for all $c_j\in C_j$ and for any $p\in V$,

$$\int_{\mathcal{D}}D(p)\sigma^*_{ij}(\mathrm{d}D|c_j)=D_0^{ij}(p). \qquad (91)$$

Moreover, for G_j-almost all $c_j\in C_j$,

$$\int_{\mathcal{D}}(\overline{\boldsymbol{p}}_D(c_j)-\phi_{G_j}(c_j))D(\overline{\boldsymbol{p}}_D(c))\sigma^*_{ij}(\mathrm{d}D|c_j)=\int_V(v-\phi_{G_j}(c_j))^+D_0^{ij}(\mathrm{d}v). \qquad (92)$$

Furthermore, for Lebesgue-almost all $c_j\in C_j$,

$$\int_{\mathcal{D}}D(\overline{\boldsymbol{p}}_D(c_j))\sigma^*_{ij}(\mathrm{d}D|c_j)=D_0^{ij}(\phi_{G_j}(c_j)), \qquad (93)$$

and for all $c_j,c_j'\in C_j$ with $c_j<c_j'$, then

$$\int_{c_j}^{c_j'}\left(\int_{\mathcal{D}}D(\overline{\boldsymbol{p}}_D(z))\sigma^*_{ij}(\mathrm{d}D|c_j')\right)\mathrm{d}z\leq\int_{c_j}^{c_j'}D_0^{ij}(\phi_{G_j}(z))\,\mathrm{d}z, \qquad (94)$$

while for Lebesgue-almost all $c_j,c_j'\in C_j$ with $c_j'<c_j$,

$$\int_{\mathcal{D}}D(\overline{\boldsymbol{p}}_D(c_j))\sigma^*_{ij}(\mathrm{d}D|c_j')=D_0^{ij}(\phi_{G_j}(c_j')). \qquad (95)$$

Now let q^* be defined as

$$q^*_{ij}(c):=\frac{1}{|W_j(c_j)|}\mathbf{1}\left\{\int_V(v-\phi_{G_j}(c_j))^+D_0^{ij}(\mathrm{d}v)=\max_{i'\in\mathcal{I}}\int_V(v-\phi_{G_j}(c_j))^+D_0^{i'j}(\mathrm{d}v)\right\},$$

where

$$W_j(c_j):=\left\{i\in\mathcal{I}\,\middle|\,\int_V(v-\phi_{G_j}(c_j))^+D_0^{ij}(\mathrm{d}v)=\max_{i'\in\mathcal{I}}\int_V(v-\phi_{G_j}(c_j))^+D_0^{i'j}(\mathrm{d}v)\right\}.$$

Then the data broker's revenue under σ^* and q^* attains the upper $\bar{R}$. Furthermore, notice that since the function

$$z \mapsto \int_{\{v \geq z\}} (v-z) D(\mathrm{d}v)$$

is nonincreasing for any $D \in \mathcal{D}$ and since ϕ_{G_j} is nondecreasing, q^*_{ij} is nonincreasing for each $i \in \mathcal{I}$. As a result, for each $j \in \mathcal{J}$, for any $c_j, c'_j \in C_j$

$$\begin{aligned}
&\int_{c_j}^{c'_j} \left(\mathbb{E}_{c_{-j}} \left[\sum_{i \in \mathcal{I}} \int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(z)) \sigma^*_{ij}(\mathrm{d}D|z, c_{-j}) q^*_{ij}(z, c_{-j}) \right] \right) \mathrm{d}z \\
&\quad - \int_{c_j}^{c'_j} \left(\mathbb{E}_{c_{-j}} \left[\sum_{i \in \mathcal{I}} \int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(z)) \sigma^*_{ij}(\mathrm{d}D|c'_j, c_{-j}) \right] q^*_{ij}(c'_j, c_{-j}) \right) \mathrm{d}z \\
&= \int_{c_j}^{c'_j} \left(\int_{\mathcal{D}} \sum_{i \in \mathcal{I}} D(\overline{\boldsymbol{p}}_D(z)) \sigma^*_{ij}(\mathrm{d}D|z) q^*_{ij}(z) - \int_{\mathcal{D}} \sum_{i \in \mathcal{I}} D(\overline{\boldsymbol{p}}_D(z)) \sigma^*_{ij}(\mathrm{d}D|c'_j) q^*_{ij}(c'_j) \right) \mathrm{d}z \\
&\geq \int_{c_j}^{c'_j} \sum_{i \in \mathcal{I}} \left(D_0^{ij}(\phi_{G_j}(z)) q^*_{ij}(z) - D_0^{ij}(\phi_{G_j}(c'_j)) q^*_{ij}(c'_j) \right) \mathrm{d}z \\
&\geq 0,
\end{aligned} \tag{96}$$

where the equality follows from the fact that σ^*_{ij} and q^*_{ij} do not depend on c_{-j}, the first inequality follows from (93), (94), and (95); and the last inequality follows from monotonicity of ϕ_{G_j} and q^*_{ij} for each $i \in \mathcal{I}$.

Finally, notice that since for each $j \in \mathcal{J}$, $\{D_0^{ij}\}_{i \in \mathcal{I}}$ is ordered by pointwise dominance, for any $c_j \in C_j$, any $i, i' \in \mathcal{I}$

$$\begin{aligned}
&\int_V (v - \phi_{G_j}(c_j))^+ D_0^{ij}(\mathrm{d}v) \geq \int_V (v - \phi_{G_j}(c_j))^+ D_0^{i'j}(\mathrm{d}v) \\
&\iff m^{D_0^{ij}} \geq_{\mathrm{FOSD}} m^{D_0^{i'j}} \\
&\iff D_0^{ij} \geq D_0^{i'j}.
\end{aligned} \tag{97}$$

As a result, for any $j \in \mathcal{J}$, for any $c_j \in C_j$,

$$\begin{aligned}
&\int_{c_j}^{\bar{c}_j} \left(\mathbb{E}_{c_{-j}} \left[\sum_{i \in \mathcal{I}} \int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(z)) \sigma^*_{ij}(\mathrm{d}D|z, c_{-j}) q^*_{ij}(z, c_{-j}) \right] \right) \mathrm{d}z \\
&= \int_{c_j}^{\bar{c}_j} \sum_{i \in \mathcal{I}} \left(\int_{\mathcal{D}} D(\overline{\boldsymbol{p}}_D(z)) \sigma^*_{ij}(\mathrm{d}D|z) q^*_{ij}(z) \right) \mathrm{d}z \\
&= \int_{c_j}^{\bar{c}_j} \sum_{i \in \mathcal{I}} D_0^{ij}(\phi_{G_j}(z)) q^*_{ij}(z) \, \mathrm{d}z \\
&= \int_{c_j}^{\bar{c}_j} \sum_{i \in W_j(z)} D_0^{ij}(\phi_{G_j}(z)) \frac{1}{|W_j(z)|} \, \mathrm{d}z \\
&\geq \int_{c_j}^{\bar{c}_j} \sum_{i \in \mathcal{I}} D_0^{ij}(\phi_{G_j}(c_j)) \frac{1}{I} \, \mathrm{d}z \\
&\geq \int_{c_j}^{\bar{c}_j} D_0^j(\overline{\boldsymbol{p}}_{D_0^j}(z)) \, \mathrm{d}z,
\end{aligned} \tag{98}$$

where the first equality follows from the fact that σ^*_{ij} and q^*_{ij} do not depend on c_{-j} for all $i \in \mathcal{I}$, the second equality follows from (93), the third equality is from the definition of $\{q^*_{ij}\}$, the fourth equality follows from the definition of q^* and from (97), and the last equality follows from that fact that $\phi_{G_j} \leq \overline{\boldsymbol{p}}_{D_0^j}$ for all $j \in \mathcal{J}$.

Together, from (96) and (98), there exists transfers $\{\tau^*_j\}_{j\in\mathcal{J}}$ such that (σ^*, τ^*, q^*) is incentive compatible. Moreover, for each $j \in \mathcal{J}$, by taking $\tau^*_j(\bar{c}_j)$ as $\int_{\{v \geq \bar{c}_j\}} \sum_{i\in\mathcal{I}}(v - \bar{c}_j)D_0^{ij}(\mathrm{d}v)q^*_{ij}(\bar{c}_j) - \bar{\pi}_j$, together with (92), (98) and Lemma 1, the mechanism (σ^*, τ^*, q^*) is indeed incentive feasible and attains the upper bound $\bar{R}$.

Finally, it remains to show that the producers' gross expected profit and the allocation of the product under any optimal mechanism are the same as those under the price-controlling data broker's optimal mechanism. Since the optimal mechanism (σ^*, τ^*, q^*) constructed above attains the upper bound $\bar{R}$, all the inequalities in (90) are binding, by exactly the same arguments as in the proof of Proposition 2 and Theorem 7 and by noticing that any optimal mechanism (σ, τ, q) for the data broker of any optimal mechanism $(\sigma, \tau, q, \boldsymbol{\gamma})$ for the price-controlling data broker must entail

$$q_{ij}(c) > 0 \iff \int_V (v - \phi_{G_j}(c_j))D_0^{ij}(\mathrm{d}v) = \max_{i'\in\mathcal{I}} \int_V (v - \phi_{G_j}(c_j))D_0^{i'j}(\mathrm{d}v),$$

under any optimal mechanism of either the data broker or the price-controlling data broker, the allocation of the product must be such that for each product j, all the consumers in group $i(j)$ buys product j by paying their values and the rest of the consumers do not buy, where $i(j)$ is the group that prefers j the most (i.e. $i(j)$ is such that $D_0^{i(j)j} \geq D_0^{ij}$ for all $i \in \mathcal{I}$), while each producer $j \in \mathcal{J}$ must have expected profit

$$\max_{i\in\mathcal{I}} \int_{C_j} \int_{\{v \geq \phi_{G_j}(c_j)\}} (v - c_j)D_0^{ij}(\mathrm{d}v),$$

which are exactly the allocation and the gross profit producer $j \in \mathcal{J}$ earns when facing the price-controlling data broker, respectively. This completes the proof. ■